Andrews & Hunt Digital Image Restoration
Bracewell Two Dimensional Imaging
Brigham The Fast Fourier Transform and Its Applications
Burdic Underwater Acoustic System Analysis, 2/E
Cohen Time Frequency Analysis
Crochiere & Rabiner Multirate Digital Signal Processing
Dudgeon & Mersereau Multidimensional Digital Signal Processing
Haykin Advances in Spectrum Analysis and Array Processing, Vols. I, II, & III
Haykin, ed. Array Signal Processing
Johnson & Dudgeon Array Signal Processing
Kay Fundamentals of Statistical Signal Processing
Kay Modern Spectral Estimation
Lim Two-Dimensional Signal and Image Processing
Lim & Oppenheim, eds. Advanced Topics in Signal Processing
Marple Digital Spectral Analysis with Applications
McClellan & Rader Number Theory in Digital Signal Processing
Nikias & Petropulu Higher-Order Spectra Analysis
Oppenheim & Nawab Symbolic and Knowledge-Based Signal Processing
Oppenheim, Willsky, with Young Signals and Systems
Oppenheim & Schafer Digital Signal Processing
Oppenheim & Schafer Discrete-Time Signal Processing
Picinbono Random Signals and Systems
Rabiner & Juang Fundamentals of Speech Recognition
Rabiner & Schafer Digital Processing of Speech Signals
Stearns & David Signal Processing Algorithms in Fortran and C
Therrien Discrete Random Signals and Statistical Signal Processing
Tribolet Seismic Applications of Homomorphic Signal Processing
Vaidyanathan Multirate Systems and Filter Banks
Vetterli & Kovacevic Wavelets and Subband Coding
Widrow & Stearns Adaptive Signal Processing
Wornell Signal Processing with Fractals

Signal Processing with Fractals:
A Wavelet-Based Approach

Gregory Wornell
Massachusetts Institute of Technology

UNIVERSIDAD NACIONAL MEXICO
NM135693
X090916
ENDCol.

For book and bookstore information
http://www.prenhall.com

Prentice Hall PTR
Upper Saddle River, NJ 07458
Contents

Preface x

1 Introduction and Overview 1

2 Wavelet Transformations 8

2.1 Introduction 8

2.2 Wavelet Bases 9

2.3 Orthonormal Wavelet Bases 11

2.3.1 An Octave-Band Filter Bank Interpretation 13

2.3.2 Multiresolution Signal Analysis Interpretation 16

2.3.3 Discrete Wavelet Transform 21

2.3.4 Finite Data Length and Resolution Effects 23

2.3.5 Orthonormal Wavelet Basis Constructions 25

2.3.6 Examples 26

2.3.7 Nondyadic Orthonormal Wavelet Bases 27

2.4 Summary 28

3 Statistically Self-Similar Signals 30

3.1 Introduction 30

3.2 1/f Processes 32

3.2.1 Fractional Brownian Motion and Fractional Gaussian Noise 36
Preface

In recent years, the mathematics of fractal geometry has generated much excitement within the engineering community among those seeking broad new and more realistic classes of models for wide-ranging applications. This optimism has arisen out of the perspective that many natural and man-made phenomena around which we engineer our world are much better described through an inherently irregular "fractal" geometry than by the traditional regular Euclidean geometry. However, despite the apparent promise of fractal geometry in providing useful solutions to important engineering problems, progress in applying these models in such applications has been slower than expected. To a large degree, this has been due to the lack of an adequate set of convenient and efficient engineering-oriented mathematical tools for exploiting these ideas.

From this perspective, the recent emergence of powerful multiscale signal representations in general and wavelet basis representations in particular has been particularly timely. Indeed, out of this theory arise highly natural and extremely useful representations for a variety of important fractal phenomena. This book presents both the development of these new techniques as well as their application to a number of fundamental problems of interest to signal processing and communications engineers.

In particular, this book develops a unified, wavelet-based framework for efficiently synthesizing, analyzing, and processing several broad classes of fractal signals. For example, efficient and practical algorithms for solving some important problems of optimal estimation, detection and classification involving fractals are developed using this framework. As another example, novel and practical signal processing techniques for exploiting fractal signals as information-bearing waveforms in efficient communication systems are also developed.

In preparing this monograph, there has been an effort to make this material as accessible as possible to graduate students and practicing professionals alike. In particular, no prior familiarity with either fractal geometry or wavelets is assumed of the reader. Indeed, Chapter 2 is a fully self-contained primer on the relevant wavelet theory. However, to keep the treatment compact, the reader is assumed to possess a basic familiarity with continuous- and discrete-time signals and systems, with stochastic processes, detection and estimation theory, and with the geometry of linear vector spaces and linear algebra. Wherever possible, concepts and intuition are emphasized over mathematical rigor. Nevertheless, unavoidably some measure of that ill-defined but important quantity "mathematical sophistication" is also assumed of the reader.

As a final remark, it should be emphasized that no attempt has been made to describe many of the exciting parallel developments taking place within this active field of research. While I have tried whenever possible to point out those developments most closely related to the material covered, this monograph should not be interpreted as a comprehensive treatise. In fact, rather than a retrospective on a mature topic, it is hoped that this book will serve as a catalyst, stimulating further development in both the theory and applications of the exciting, powerful, and increasingly important ideas in this area.

There are many people who contributed to this book becoming a reality, and to them I am extremely grateful. Alan Oppenheim contributed generously to the development of the ideas in the original work, and strongly encouraged and supported the book. Alan Willsky and William Siebert also provided valuable technical input and suggestions during the development of this material. Henrique Malvar, Jelena Kovacevic, and Jerome Shapiro all read the complete manuscript very carefully and provided a great deal of helpful and detailed feedback. Warren Lam and Haralabos Papadopoulos also did an exceptional job proofreading and helping to debug the manuscript. Jon Sjogren and his program at the Air Force Office of Scientific Research played a critical role in supporting and encouraging the research out of which this book grew. And finally, Karen Gettrnan at Prentice-Hall was instrumental in steering the project to fruition.

Gregory W. Wornell
Cambridge, Massachusetts
Signal processing plays a central role of a truly enormous range of modern electronic systems. These include, for example, voice, data, and video communication and storage systems; medical imaging and diagnostic systems; radar, sonar and satellite remote sensing systems; and automated manufacturing systems. Signal processing algorithms required for these kinds of applications take many forms. Efficient speech compression and recognition systems, for example, require sophisticated signal analysis algorithms. By contrast, high-speed modems for data communication require not only powerful signal detection algorithms at the receiver, but efficient signal synthesis algorithms at the transmitter. And several other kinds of algorithms— including signal restoration, enhancement, and manipulation—are also important in diverse applications.

Whenever we construct algorithms for signal synthesis, analysis or processing, we invariably exploit some model for the signals of interest. In some instances this model is implicit. For example, spline-based strategies are often used in signal interpolation problems even when there is no quantitative model for the signal. In such cases, there is an implicit assumption that the signal is smooth at least in some qualitative sense. In other instances the model is explicit. For example, bandlimited interpolation algorithms are designed for a specific and well-defined class of signals whose frequency content is zero except within some known bandwidth.

The signal model may arise out of a deterministic formulation (e.g., a sinusoid of possibly unknown frequency and/or phase), or a stochastic one (e.g., a first-order autoregressive process). It may encompass a particularly
broad class of signals, or a relatively narrow one. The class of stationary random processes, for instance, is much larger than the class of first-order autoregressive processes. Not surprisingly, better performance can be expected of algorithms tuned to more narrowly defined classes of signals. This of course inevitably comes at the expense of robustness and flexibility: we cannot in general expect systems to behave well on signals for which they have not been designed.

Choosing a signal model for a particular application involves many factors in addition to those cited above. For example, naturally we seek models that capture the important characteristics of the physical signals of interest as closely as possible. On the other hand, an overly complex model can lead to an essentially intractable framework for algorithm development. Hence, the art and science of signal model selection invariably involves a compromise.

Much of traditional signal processing has relied upon a relatively small class of models for use in wide-ranging applications. These models include, for example, periodic and bandlimited signals. They also include important classes of stationary random signals derived from filtered white noise processes such as the autoregressive moving-average (ARMA) models. And certain cyclostationary signals turn out to be important models for many signal processing applications involving digital communications.

It is worth noting that as a whole these models have tended to arise out of signal processing's deep traditional foundation in linear time-invariant (LTI) system theory [1] [2] [3]. For example, periodic signals are characterized by a form of translational invariance: the associated waveforms are invariant to time translations by multiples of the period. Likewise, stationary and cyclostationary signals are also characterized by a form of translational invariance: the statistics of these processes are invariant to appropriate time translations of the underlying waveforms. With the close connection to LTI system theory, it is not surprising that the Fourier transform plays such a key role in the analysis and manipulation of these models, nor that fast Fourier transform (FFT) plays such a key role in the implementation of algorithms based on such models.

Interestingly, even the most sophisticated models that are used in signal processing applications typically have evolved out of useful extensions of an LTI framework. For example, the time-varying signal models required for many applications often developed out of such extensions. As a consequence windowed Fourier transform generalizations such as the short-time Fourier transform (STFT) frequently play an important role in such applications. Even state-space formulations, which can in principle accommodate a broad range of behaviors, are firmly rooted in a time-invariant perspective.

However, a strong theme that runs throughout this book is that there are many physical signals whose key characteristics are fundamentally different than can be produced by many of these traditional models and such generalizations. In particular, there are many signals whose defining characteristic is their invariance not to translation but rather to scale. For example, in the stochastic case, we mean that the statistics of the process do not change when we stretch or shrink the time axis. Both qualitatively and quantitatively the process lacks a characteristic scale: behavior of the process on short time scales is the same as it is on long time scales. An example of a process with this behavior is depicted in Fig. 1.1.

Figure 1.1. A sample waveform from a statistically scale-invariant random process, depicted on three different scales.
Typically, an important consequence of this scale-invariance is that the resulting waveforms are fractal and highly irregular in appearance [4]. In general, fractals are geometric objects having nondegenerate structure on all temporal or spatial scales. For the fractals most frequently of interest, the structure on different scales is related. The fractal processes on which we focus in particular have the property that this structure at different scales is similar. This, in turn, has the important consequence that even well-separated samples of the process end up being strongly correlated.

Collectively, these features are, of course, extremely common in many types of physical signals, imagery, and other natural phenomena. In fact, as we discuss in Chapter 3, 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 and data traffic on communication channels, in the bronchi of the human lung, and even in fluctuations of the stock market.

It is worth emphasizing that random process models traditionally used in signal processing, by contrast, typically do not exhibit this kind of behavior. Fig. 1.2, for example, depicts the behavior of a simple, first-order autoregressive process on different time scales. Hence, in order to use fractal signal models with statistical scale-invariance characteristics in applications, we require techniques for detecting, identifying, classifying, and estimating such signals in the presence of both other signals and various forms of additional distortion. In Chapter 4, we therefore develop a variety of such signal processing algorithms. As an example, we develop robust and efficient algorithms for estimating the fractal dimension of signals from noisy measurements, and for optimally restoring such signals.

In the deterministic case, scale-invariance means that the actual waveform is repeatedly embedded within itself, so that temporal dilations and compressions simply reproduce the signal. An example of a "self-similar" waveform of this type is depicted in Fig. 1.3 on different time scales. As we develop in Chapter 6, such signals have a potentially important role to play as modulating waveforms in, for example, secure communication applications, where they provide a novel and powerful form of transmission diversity. In particular, the resulting waveforms have the property that the transmitted information can be reliably recovered given either severe time- or band-limiting in the communication channel. In light of the geometric properties of the resulting transmitted waveforms, we refer to this diversity strategy as "fractal modulation."

At one time it was believed that relatively complicated signal models were necessary to produce the kind of special behavior we have described. However, as will become apparent in subsequent chapters, while fundamentally new model structures are required, they are in fact no more complex than (and in many respects at least as tractable as) many traditional signal processing models. In particular, remarkably simple characterizations that are useful both conceptually and analytically are developed in Chapter 3 for the case of statistically scale-invariant signals, and in Chapter 5 for the case of deterministically scale-invariant signals.

In addition, while the Fourier transform plays a central role in the analysis and manipulation of both statistically and deterministically translation-invariant signals, as we will see it is the wavelet transform that plays an analogous role for the kinds of scale-invariant signal models that are the focus of this book. Wavelet representations are expansions in which the basis
signals are all dilations and translations of a suitable prototype function. In Chapter 3, for example, we develop the role of such wavelet representations as Karhunen-Loève type expansions for stochastically scale-invariant random processes: when projected onto wavelet bases, the resulting expansion coefficients are effectively uncorrelated. In Chapter 5, we show how wavelet bases can be used to construct orthonormal bases of self-similar waveforms that provide efficient representations for deterministically scale-invariant waveforms.

Yet wavelet representations are not merely a conceptual tool in the manipulation of fractal signals. In fact, just as the FFT plays an important role in practical algorithms for processing translation-invariant signals, it is the discrete wavelet transform (DWT) that is frequently a key component of practical algorithms for processing scale-invariant signals. In Chapter 4, for example, we see how fast algorithms for estimating the fractal dimension of signals rely heavily on the DWT. Similarly, in Chapter 6, we see how the DWT algorithm is critical to efficient implementations of fractal modulation transmitters and receivers.

Finally, in Chapter 7 we explore some system theoretic foundations for the classes of signals developed in earlier chapters. We discuss the duality between time-invariant and scale-invariant systems, and show how Mellin, Laplace, and wavelet transforms arise rather naturally in the representation of the latter class of systems. In the process, some interesting and unifying themes are introduced.

Given the central role that the wavelet transform plays in the efficient representation of scale-invariant signals, we begin with a self-contained overview of the relevant aspects of wavelet theory in Chapter 2.
Wavelet Transformations

2.1 Introduction

Wavelet transformations play a central role in the study of self-similar signals and systems. As we will 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 of 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 for self-similar signals are also of great practical significance.

In terms of history, the theory of wavelet transformations dates back to the work of Grossmann and Morlet [5], and was motivated by applications in seismic data analysis [6]. This inspired much subsequent work by several individuals over the next several years on a general mathematical theory. For example, several key results in the theory of nonorthogonal wavelet expansions are described by Daubechies [7] [8]. In this book, however, we are 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 [9] and Mallat [10] [11]. However, it was Daubechies who introduced the first highly practical families of orthonormal wavelet bases in her landmark paper [12].

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 [13],
- multigrid methods in the solution of partial-differential and integral equations [14],
- spectrogram methods in speech recognition [15],
- progressive transmission algorithms and embedded coding in communications [16] [17] [18], and
- multirate filtering algorithms in digital audio [19] [20], speech and image coding [21] [22], voice scrambling [23], frequency division data multiplexing [24], and time/frequency division data cross-multiplexing [20].

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 [25], and quadrature mirror, conjugate quadrature, and other related multirate filter banks [20]. Vetterli and Kovacevic [22], for example, explore such connections in detail.

This chapter is designed as a self-contained overview of wavelet transformations in general and of orthonormal wavelet transformations in particular. Its primary purpose is to establish some notational conventions for wavelets and to summarize the key results from wavelet theory we exploit in the applications in subsequent chapters. In introducing the theory of wavelet transformations, we adopt a rather tutorial style and emphasize a strongly signal processing perspective. While a number of excellent introductions to wavelet theory can be found in the literature (see, e.g., [7] [11] [26] [27]) we stress that the one presented here emphasizes a perspective that is particularly important in light of the applications that are the focus of this book.

2.2 Wavelet Bases

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

\[ x(t) \leftrightarrow X^\delta \]

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.
2.3 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 more fully developed. In Section 2.3.7, 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 \( x(t) \)

\[ x(t) \leftrightarrow x_n^m \]

can be described in terms of the synthesis/analysis equations\(^1\)

\[ x(t) = W_d^{-1} \{ x_n^m \} = \sum_m \sum_n x_n^m \psi_n^m(t) \quad (2.5a) \]

\[ x_n^m = W_d \{ x(t) \} = \int_{-\infty}^{\infty} x(t) \psi_n^m(t) dt \quad (2.5b) \]

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 (or "mother" 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 refer on numerous occasions throughout the book, is that derived from the ideal bandpass wavelet, which we specifically denote by \( \psi^*(t) \). This wavelet is the impulse response of an ideal bandpass filter with frequency response

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

\[ x(t) = W^{-1} \{ X_n^p \} = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X_n^p \psi_n^p(t) \mu^{-2} d\mu d\nu. \quad (2.4) \]

Under certain circumstances, it is also possible to reconstruct \( x(t) \) solely from samples of \( X_n^p \) on a hyperbolic lattice defined by

\[ \mu = a^{-m}, \quad \nu = nb^{-m} \]

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 corresponding countably infinite collection of functions \( \psi_n^m(t) \) is said to constitute a frame. A general theory and some iterative reconstruction algorithms are presented by Daubechies \cite{7} \cite{8}. 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.

\(^1\)For notational brevity, summations are over all integers, i.e., unless otherwise specified.

\(^2\)A detailed treatment of the notation, concepts, and geometry of vector spaces that provides a background for this chapter can be found in, for example, Naylor and Sell \cite{28} or Reed and Simon \cite{29}.
Wavelet Transformations

constitute an orthonormal set we get the Poisson formula

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

and, in turn, the bound

$$|\psi(\omega)| \leq 1.$$  \hfill (2.8a)

Moreover, from (2.3) we have immediately

$$\psi(0) = 0.$$  \hfill (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 \( f(t) \) is said to be \( R \)-th-order regular if its Fourier transform \( \mathcal{F}(\omega) \) decays according to\(^3\)

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

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

$$\psi(\omega) \sim \mathcal{O}(|\omega|^{-1}), \quad |\omega| \to \infty.$$  \hfill (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 will see later, in terms of a multiresolution signal analysis. While we 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 [30]

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

generates a perfectly valid orthonormal wavelet basis, but does not correspond to a multiresolution analysis.

3The order notation \( \mathcal{O}(\cdot) \) has the following interpretation: if

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

then

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

2.3.1 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.,

$$x[n] = \{x(t) * \psi^n(\tau - t)\}_{\tau=-N_\text{max}}^{\text{N}_\text{max}}.$$  

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 [25], 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, perfect reconstruction and orthogonality are achieved due to the characteristics of the phase in the filters, which leads to cancellation of aliasing effects in the reconstruction. However, it is important to emphasize that 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, \((x^m_n)\) represents an estimate of the energy of the signal \(x(t)\) in the vicinity of \(t \approx 2^{-m}n\), and for a band of frequencies in the neighborhood of \(\omega \approx 2^{-m}\pi\). 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 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 \(x^m_n\) is modulated onto the corresponding wavelet dilate \(\psi^m_0(t)\) at rate \(2^m\). For the case of the ideal bandpass wavelet, this corresponds to modulating each such sequence \(x^m_n\) 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

\[
\sum_m |\Psi(2^{-m}\omega)|^2 = 1
\]

(2.9)

valid for all orthonormal wavelet bases and any \(\omega \neq 0\). To see this, 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 re-synthesized 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

$$X(\omega) = X_0(\omega) + \sum_{k \neq 0} X(\omega - 2\pi k 2^m) \psi(2^{-m}\omega) \psi^*(2^{-m}\omega - 2\pi k).$$

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 [9] [11] [12]. This perspective, which we consider next, provides a number of rich new insights into orthonormal wavelet bases.

### 2.3.2 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 [31] and multigrid methods used in the solution of boundary value problems [14]. 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.

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 restricting our attention to the vector space of finite-energy signals $V = L^2(\mathbb{R})$. A multiresolution signal analysis is then defined as a decomposition of this signal space $V$ into a sequence of subspaces $V_0, V_1, V_2, \ldots$ such that each $V_m$ contains signal approximations at a resolution $2^m$. Associated with each $V_m$ is a linear operator $A_m\{\cdot\}$ that defines projections from anywhere in $V$ onto $V_m$. That is, for each signal $x(t) \in V$, the projection $A_m\{x(t)\} \in V_m$ defines the closest signal of resolution $2^m$ to $x(t)$,

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

Central to the concept of multiresolution signal 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

$$V_m \subset V_{m+1}, \quad (2.10a)$$

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 $2^m$, i.e.,

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

The relations

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

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

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.,
\[
\lim_{m \to \infty} A_m \{ x(t) \} = x(t),
\]
\[
\lim_{m \to -\infty} A_m \{ x(t) \} = 0.
\]

An additional relation is required to fully define the notion of resolution: signals in \( V_m \) must be characterized by \( 2^m \) samples per unit length. Mathematically, this can be interpreted as requiring that there exist a one-to-one correspondence or “isometry” between each subspace of signals \( V_m \) and the vector space of finite-energy sequences \( \mathbb{I} = l^2(\mathbb{Z}) \)

\[ V_m \overset{isom}{\leftrightarrow} \mathbb{I} \tag{2.10d} \]

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

\[ x(t) \in V_m \Rightarrow \varphi_m \{ x(t) \}_{n=2^m n} \in \mathbb{I} \tag{2.10e} \]

where \( \varphi_m \{ \cdot \} \) 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 convenient structure. The first is a translation-invariance constraint, viz.,

\[ x(t) \in V_m \Leftrightarrow x(t - 2^m n) \in V_m \tag{2.10f} \]

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 leads to the translational relationships among basis functions in the corresponding wavelet expansion. The second is a scale-invariance constraint

\[ x(t) \in V_m \Leftrightarrow x(2t) \in V_{m+1} \tag{2.10g} \]

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

It can be shown [30] that every such multiresolution analysis, 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 \( A_m \{ \cdot \} \) for each of these subspaces. In particular, for each \( m \),

\[ \ldots, \phi_{-1}^m(t), \phi_0^m(t), \phi_1^m(t), \phi_2^m(t), \ldots \]

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 another, i.e.,

\[ \phi_n^m(t) = 2^{m/2} \phi(2^m t - n). \tag{2.11} \]

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

\[ A_m \{ x(t) \} = \sum_n a_n^m \phi_n^m(t) \tag{2.12} \]

with the coefficients \( a_n^m \) computed according to the individual projections

\[ a_n^m = \int_{-\infty}^{\infty} x(t) \phi_n^m(t) \, dt. \tag{2.13} \]

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

\[ a_n^m = \{ x(t) \ast \phi_n^m(-t) \}_{2^m n}. \tag{2.15} \]

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 \( \phi(t) \), whose Fourier transform is the frequency response of an ideal lowpass filter, i.e.,

\[ \Phi(\omega) = \begin{cases} 1 & |\omega| \leq \pi \\ 0 & |\omega| > \pi \end{cases}. \tag{2.16} \]

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 [32]. In particular, \( \phi(t) \) in (2.15) plays the role of an anti-aliasing filter [3], 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 $\phi_n^m(t)$, are unrealizable. For this reason, this analysis is primarily of interest for its conceptual rather than practical value.

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 $O_m$

$$\ldots, \psi_m^m(t), \psi_0^m(t), \psi^m_1(t), \psi^m_2(t), \ldots$$

where $\psi^m_0(t)$ 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 $D_m\{\cdot\}$ 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 $m$ correspond 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.,

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

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 $M \to \infty$ 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

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

Our interpretation of an orthonormal wavelet basis as a multiresolution signal analysis is then complete.

### 2.3.3 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 (2.3.1), the analysis can be reformulated 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

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

$$g[n] = \int_{-\infty}^{\infty} \phi_0^1(t) \phi_n^0(t) \, dt. \quad (2.20b)$$

Typically, $h[n]$ and $g[n]$ have discrete-time Fourier transforms [3]

$$H(\omega) = \sum_n h[n] e^{-j\omega n}$$

$$G(\omega) = \sum_n g[n] e^{-j\omega n}$$

that have roughly halfband lowpass and highpass characteristics, respectively. In fact, for the particular case of bandlimited multiresolution signal analysis, the corresponding filters, which we distinguish using the notation
$h[n]$ and $g[n]$, are ideal lowpass and highpass filters; specifically

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

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

More generally, as we will see, the filters $h[n]$ and $g[n]$ form a conjugate quadrature filter pair.

The analysis algorithm is structured as follows. Given a signal $x(t) \in V$ from which we would like to extract $X$: for $m < M$, 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

$$a_{n}^{m+1} = \sum_{l} h[l - 2n] a_{l}^{m+1}$$  \hspace{1cm} (2.21a)

$$x_{n}^{m} = \sum_{l} g[l - 2n] a_{l}^{m+1}$$  \hspace{1cm} (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

$$a_{n}^{m} = \sum_{l} \left( h[l - 2n] a_{l}^{m} + g[l - 2n] x_{l}^{m} \right)$$  \hspace{1cm} (2.21c)

to compute the coefficients $a_{n}^{m}$ of successively finer scale approximations until level $m = M$ is reached, after which $x_{n}^{M+1}$ 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. 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 $O(\log L)$ per input sample [33]. However, as also discussed in [33] this figure can be somewhat misleading as there are many subtle issues associated with measuring complexity of the algorithm; a thorough treatment can be found in, e.g., Vetterli and Kovačević [22].

### Sec. 2.3 Orthonormal Wavelet Bases

#### 2.3.4 Finite Data Length and Resolution Effects

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

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

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.,

$$x[n] = a_{n}^{M+1} = \{ \phi_{M+1}^{M+1}(t) \ast x(t) \} |_{t = 2^{-(M+1)n}}$$

for some $M$, this cannot always be justified. Nevertheless, if the signal $x(t)$ was processed by 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 $\phi_{M+1}^{M+1}(t)$ 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 \( x(t) \) corresponding to \( m = 1, 2, \ldots, M \)

\[
\begin{align*}
    n &= 0, 1, \ldots, N_0 2^{m-1} - 1 \\
\end{align*}
\]

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. This is the method we adopt in the simulations described in this book.

Sec. 2.3 Orthonormal Wavelet Bases

2.3.5 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 the orthogonality constraints between the \( \{ \psi_n^m(t) \} \) and \( \{ \phi_n^m(t) \} \), one can show [30] that \( h[n] \) and \( g[n] \) must be related by

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

which, expressed in the frequency domain, is

\[
G(\omega) = e^{-j\omega} H^*(\omega + \pi).
\]

Furthermore, orthonormality of the \( \{ \phi_n^m(t) \} \) requires that \( H(\omega) \) satisfy

\[
|H(0)|^2 = 2
\]

\[
|H(\omega)|^2 + |H(\omega + \pi)|^2 = 2.
\]

Filter pairs that satisfy both (2.22) and (2.23) are termed conjugate quadrature filters (CQFs), and have been discussed extensively in the signal processing literature; see, e.g., Vaidyanathan [20].

We 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, even more significantly, we note that \( h[n] \) alone is also sufficient to fully characterize a wavelet basis through a multiresolution analysis. Indeed, given \( h[n] \), the dilation equation\(^4\) (A.1a) can be solved for the corresponding scaling function \( \phi(t) \). In particular, \( \phi(t) \) has Fourier transform

\[
\Phi(\omega) = \prod_{m=1}^{M} \left[ 2^{-1/2} H(2^{-m} \omega) \right]
\]

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

\(^4\)For a further discussion of dilation equations, see, e.g., Strang [26].
both (2.23) and to have a Fourier transform \( H(\omega) \) with \( R \) zeros at \( \omega = \pi \), i.e.,
\[ H^{(r)}(\pi) = 0, \quad r = 0, 1, \ldots, R - 1 \]
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:
\[ \int_{-\infty}^{\infty} t^r \varphi(t) \, dt = (j)^r \Phi^{(r)}(0) = 0, \quad r = 0, 1, \ldots, R - 1. \]  
(2.25)
It is important to note, however, that the vanishing moment condition (2.25), while convenient and sufficient, is not necessary for regularity. For a more detailed discussion of necessary and sufficient conditions, see, e.g., Lawton [34].

A variety of useful wavelet bases have been constructed from CQF pair formulations of this type. In fact, this approach has been extremely useful in designing orthonormal wavelets with compact support, i.e., wavelets for which
\[ \varphi(t) = 0, \quad |t| > T \]
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 comprehensive development of the relationships between wavelet theory and filter bank theory can be found in, e.g., Vetterli and Kovačević [22].

### 2.3.6 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 \( \varphi(t) \) decays only like \( 1/t \) for large \( t \), and the CQF filters \( h[n] \) and \( g[n] \) decay only like \( 1/n \) for large \( n \). More serious still, this basis is unrealizable.

At another 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
\[ \phi(t) = \begin{cases} 1 & 0 \leq t < 1 \\ 0 & \text{otherwise} \end{cases} \]
and the corresponding wavelet is
\[ \psi(t) = \begin{cases} 1 & 0 \leq t < 1/2 \\ -1 & 1/2 \leq t < 1 \\ 0 & \text{otherwise} \end{cases} \]

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 \( \omega \to \infty \).

Between these extremes lie a number of wavelet families. Consider, for example, the family of Battle-Lemarie wavelet bases [30] [12]. 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
\[ \Phi(\omega) = \frac{1}{\omega} \left[ \sum_{k} \frac{1}{(\omega + 2\pi k)^{P+1}} \right]^{-1/2}. \]

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 a Central Limit Theorem argument it is possible to show that the bandpass wavelet basis corresponds to \( P \to \infty \). In general, 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.

As another family of examples, Daubechies has designed an important class of compactly supported wavelet bases [12] 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 CQF filters \( h[n] \) and \( g[n] \) of length \( 2R \) for \( R = 1, 2, \ldots \), 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. For example, some with particularly attractive computational implementations are described in [35].

### 2.3.7 Nondyadic Orthonormal Wavelet Bases

While we have focused largely upon dyadic wavelet bases, for which the dilation and translation increments are \( a = 2 \) and \( b = 1 \), there are many other nondyadic choices. In many applications, including those within the context of this book, 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, it would be highly convenient to have the flexibility of choosing from among a family of bases corresponding to the lattice
\[ a = \frac{(L + 1)}{L}, \]
\[ b = L, \]
where \( L = 1, 2, \ldots \) is a parameter. An at least conceptually useful class of such bases arises out of a generalization of the ideal 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 (dyadic) bandpass basis. It is a straightforward exercise in analysis to verify that for each \( L \) the corresponding set \( \{ \psi_n(t) \} \), for which
\[ \psi_n(t) = \left( \frac{L + 1}{L} \right)^{m/2} \psi \left( \left( \frac{L + 1}{L} \right)^m t - nL \right), \]
is complete and orthonormal. Unfortunately, however, these basis functions have tails that decay very slowly due to their ideal frequency characteristics. This leads to comparatively poor time-domain localization properties which considerably reduces the practical value of these particular bases. Constructions for more practical nondyadic wavelet basis have been pursued by, e.g., Auscher [36], Blu [37], and Kovačević and Vetterli [38].

### 2.4 Summary

In this chapter, we presented those aspects of wavelet theory that will be important in the developments throughout the remainder of the book. Two interpretations of the orthonormal wavelet transform were discussed: one based on an octave-band filter bank framework; the other on a multiresolution signal analysis framework. Each provided rich and complementary insights into wavelet theory. Our treatment of the topic emphasized a signal processing perspective, and developed strong intuition in both the time and frequency domains.

The chapter also discussed several aspects of the implementation of the wavelet transform in practice, since these will be important when wavelets are exploited in the development of actual fractal signal processing algorithms in subsequent chapters. In particular, we discussed the discrete wavelet transform, a discrete-time algorithm for synthesizing and analyzing signals via the orthonormal wavelet transform. Related issues including the effects of finite data length and limited data resolution on the computation were also discussed. The chapter concluded with several examples of families of wavelet bases that are important both conceptually and practically.
Chapter 3

Statistically Self-Similar Signals

3.1 INTRODUCTION

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 the 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 [39]. We then introduce and study an alternative mathematical characterization for $1/f$ processes. The novelty and power of this characterization are 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 a 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 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 an appropriate 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.

Before proceeding with our development, we point out that a basic engineering background in probability, random variables, and random processes is assumed of the reader throughout this chapter. Useful treatments of this background material can be found in, e.g., Papoulis [40] or Stark and Woods [41].

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

$$x(t) \overset{P}{=} a^{-H}x(at)$$

where $P$ 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

$$M_x(t) \overset{P}{=} E \{x(t)\} = a^{-H}M_x(at)$$

$$R_x(t,s) \overset{P}{=} E \{x(t)x(s)\} = a^{-2H}R_x(at,as).$$

(3.2a)

(3.2b)

We restrict our attention to Gaussian processes, for which the two definitions are equivalent. Furthermore, we 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 autocorrelation 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 $w(t)$ by

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

(3.3)

Throughout 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 nondifferentiability 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
and extended to $t < 0$ through the convention
\[ \int_{t}^{0} \triangleq - \int_{0}^{t} \]
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
\[ x(t) = |t|^{H_{o} - 1/2} z(t) \]  
(3.5)
is self-similar with parameter $H_{o}$ for all values of $H_{o}$ is nonstationary, and has a nonstationary derivative except for $H_{o} = 1/2$. In fact, when $x(t)$ in (3.5) is filtered by virtually any nontrivial 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 so-called “1/f processes.”

### 3.2 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 form
\[ S_{x}(\omega) \sim \frac{\sigma^{2}}{|\omega|^{\gamma}} \]  
(3.6)
for some spectral parameter $\gamma$ related to $H$ according to
\[ \gamma = 2H + 1. \]  
(3.7)
Generally, the power law relationship (3.6) extends over several decades of frequency. While data length typically limits access to spectral information at lower frequencies, data resolution typically 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 [39] [42] [43] [44].

As a consequence of their inherent self-similarity, the sample paths of 1/f processes are typically fractals [4]. In general, the graphs of sample paths corresponding Riemann-Stieltjes integral is required. Nevertheless, we retain the notation $w(t)$ for conceptual convenience.

Sec. 3.2 1/f Processes

of random processes are one-dimensional curves in the plane. This we refer to as their “topological dimension.” However, fractal random processes have sample paths that are so irregular that their graphs have an “effective” dimension that exceeds their topological dimension of unity. It is this effective dimension that is usually referred to as the “fractal” dimension of the graph. However, it is important to note that the notion of fractal dimension is not uniquely defined. There are several different definitions of fractal dimension—each with subtle but important differences—from which to choose for a given application [45]. Nevertheless, regardless of the particular definition, the fractal dimension $D$ of the graph of an ordinary function typically ranges between $D = 1$ and $D = 2$. Larger values of $D$ correspond to functions whose graphs are increasingly rough in appearance and, in an appropriate sense, fill the plane in which the graph resides to a greater extent.

For 1/f processes, there is a strong relationship between the fractal dimension $D$ and the self-similarity parameter $H$ of the process. In particular, an increase in the parameter $H$ yields a decrease in the dimension $D$. This is intuitively reasonable: an increase in $H$ corresponds to an increase in $\gamma$, 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 $\gamma$. It is worth noting 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. In Section 3.2.1 we describe the quantitative relationship between a particular notion of fractal dimension $D$ and the self-similarity parameter $H$ for the class of fractional Brownian motion models for 1/f processes.

A truly enormous and tremendously varied collection of natural phenomena exhibit 1/f-type spectral behavior over many decades of frequency. A partial list includes (see, e.g., [4] [43] [46] [47] [48] [49] 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;
Statistically Self-Similar Signals

Sec. 3.2 1/f Processes

- 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 $\gamma$ 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 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 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

\[ \frac{|\tau|^{\gamma-1}}{2\Gamma(\gamma) \cos(\gamma \pi/2)} \leftrightarrow \frac{1}{|\omega|^\gamma} \]

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

\[ R_\gamma(\tau) \sim |\tau|^{-\gamma}. \]

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 that 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 [39] [51] [52] [53]. 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 the sequel.

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. Unavoidably, several mathematical subtleties arise in the development of fractional Brownian motion, making this section somewhat less accessible to the nonspecialist than others. However, while
3.2.1 Fractional Brownian Motion and Fractional Gaussian Noise

It is generally agreed [4] [39] [53] 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 [39] and promoted their use in numerous subsequent publications (see, e.g., the references in [4]). An extensive bibliographic guide to various subsequent developments of the theory, principally in the mathematics literature, is presented in Taqqu [53].

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 directly accommodate the cases $\gamma > 3$ and $\gamma < -1$, although extensions can be formulated. Furthermore, somewhat disturbingly, the models are degenerate for the cases $\gamma = -1$, $\gamma = 1$, and $\gamma = 3$.

To develop the concept, we begin by exploring the possibility of developing $1/f$ models as the result of driving stationary white Gaussian noise through a suitable linear time-invariant system. In this case, a natural choice would be the system with impulse response

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

where $u(t)$ is the unit-step function

$$u(t) = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases},$$

and where $\Gamma(\cdot)$ is the gamma function. Indeed, (3.9) has the generalized Laplace transform [54]

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

which suggests that if the input $w(t)$ has spectral density $\sigma_w^2$, the output will have a power spectrum, in some sense, of the form (3.6) where $\gamma$ is given via (3.7). As we 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 [55] 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 [56], it often is used to define the fractional integral of $w(t)$ of order $\lambda = H + 1/2 > 0$, usually denoted

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

The resulting process is well defined, satisfies $x(0) = 0$, and is statistically self-similar with parameter $H$, i.e., with $t, s, \alpha > 0$,

$$R_\alpha(t, s) = \sigma_w^2 \int_{0}^{\min(t, s)} (t - \tau)^{H-1/2} (s - \tau)^{H-1/2} \, d\tau = a^{-2H} R_\alpha(at, \alpha s) \quad (3.12)$$

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

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

while statistically self-similar, satisfying

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

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 [39].

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 Barton and Poor [57]

$$
\mathcal{D} = \frac{1}{\Gamma(H + 1/2)} \int_{-\infty}^{\infty} \left[ |t - \tau|^{H - 1/2} - |\tau|^{H - 1/2} \right] w(\tau) d\tau + \int_{0}^{t} |t - \tau|^{-H/2} w(\tau) d\tau
$$

(3.15)

for $0 < H < 1$, where $w(\tau)$ 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.

As suggested earlier, fractional Brownian motions are, in fact, fractals. Specifically, it is possible to show (see, e.g., [4] or [45]) that 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 dimension (in the Hausdorff-Besicovitch sense) given by

$$
D = 2 - H
$$

that again gives a quantitative measure of their roughness.

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

$$
R_{x}(t, s) = E[x(t)x(s)] = \frac{\sigma_{x}^{2}}{2} \left( |s|^{2H} + |t|^{2H} - |t - s|^{2H} \right)
$$

(3.16)

where

$$
\sigma_{x}^{2} = \text{var} x(1) = \frac{\Gamma(1 - 2H) \cos(\pi H)}{\pi H}
$$

(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

$$
R_{\Delta x}(\tau; \varepsilon) = \frac{\sigma_{x}^{2}}{2} \left[ \left( \frac{|\tau|}{\varepsilon} + 1 \right)^{2H} - 2 \left( \frac{|\tau|}{\varepsilon} \right)^{2H} + \left( \frac{|\tau|}{\varepsilon} - 1 \right)^{2H} \right]
$$

(3.18)

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

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

(3.19)

Letting $\varepsilon \to 0$, and defining

$$
H' = H - 1.
$$

(3.20)

we can reason from (3.15) that fractional Brownian motion has the generalized derivative [57]

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

(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

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

(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

$$
S_{x'}(\omega) = \frac{1}{|\omega|^{2H'}}.
$$

(3.23)

where

$$
g' = 2H' + 1.
$$

The preceding development suggests the conceptually useful synthesis for fractional Brownian motion depicted in Fig. 3.2. In particular, driving a linear time-invariant system with impulse response

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

with stationary white Gaussian noise $w(t)$ generates a fractional Gaussian noise $x'(t)$, from which fractional Brownian motion $x(t)$ is obtained by routine integration:

$$
x(t) = \int_{0}^{t} x'(t) dt.
$$

(3.24)

The fractional Brownian motion framework provides a useful construction for some models of $1/f$-type spectral behavior corresponding to spectral
Statistically Self-Similar Signals
Sec. 3.2 1/f Processes

In the next section, we consider a more general but nonconstructive 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. This mathematical characterization for 1/f processes was originally developed in Wornell [59] [60].

3.2.2 A Mathematical Characterization in the Frequency Domain

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, let us consider defining 1/f processes in terms of their characteristics under bandpass filtering specifically as follows.

Definition 3.1 A wide-sense statistically self-similar zero-mean random process \( x(t) \) is said to be a 1/f process if there exist \( \omega_0 \) and \( \omega_1 \) satisfying \( 0 < \omega_0 < \omega_1 < \infty \) such that when \( x(t) \) is filtered by an ideal bandpass filter with frequency response

\[
B_1(\omega) = \begin{cases} 
1 & \omega_0 < |\omega| \leq \omega_1 \\
0 & \text{otherwise} 
\end{cases}
\]

(3.24)

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

Before exploring the implications and insights that arise from this definition, it is useful to point out that choosing an ideal bandpass filter in this definition may, in fact, not be critical. It might suffice, for example, to choose any filter whose frequency response \( B(\omega) \) has sufficient decay as \( \omega \to 0 \) and \( \omega \to \infty \). Whether this leads to an equivalent definition remains an open question. Nevertheless, the use of ideal filters is certainly rather convenient. Indeed, the fundamental 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, we are able 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 these 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

\[
B(\omega) = \begin{cases} 
1 & \omega_0 < |\omega| \leq \omega_1 \\
0 & \text{otherwise} 
\end{cases}
\]

(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

$$S_y(\omega) = \begin{cases} \frac{\sigma_y^2}{|\alpha|^\gamma} & \omega_L < |\omega| \leq \omega_U \\
0 & \text{otherwise} \end{cases}$$

(3.26)

for some $\sigma_y^2 > 0$, and where the spectral exponent $\gamma$ is related to the self-similarity parameter $H$ according to $\gamma = 2H + 1$.

Of course, an important question that must be addressed concerns whether there exist any nontrivial random processes satisfying Definition 3.1. Fortunately, the answer is yes, and the following theorem constitutes an existence proof, verifying that Definition 3.1 is nondegenerate for at least some values of $\gamma$. In particular, the theorem 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 in the sense of Definition 3.1.

We remark that, based on our discussion of Section 3.2.1, an immediate corollary is that the Wiener process and stationary white Gaussian noise are also $1/f$ processes. In contrast, the Barnes-Allan process we described at the outset of Section 3.2.1 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.

Another 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 nontrivial 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 $0 < H < 1$? Recent work [61] indirectly suggests that the answer to this last question may be negative, although such a result is not explicitly proved. In effect, this paper shows 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 $\omega \to \infty$, then the definition uniquely characterizes fractional Brownian motion. However, the constraints on the filter they consider are overly restrictive to answer our specific 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 the frequency-based characterization for $1/f$ processes just described 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 will see, many of these processes retain most of the fundamental characteristics of $1/f$ processes, yet are considerably more amenable to analysis.

### 3.3 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 review two such formulations before developing a wavelet-based model for $1/f$-type behavior that is the focus of the chapter.

#### 3.3.1 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 [62]. However, it was really the seminal paper of van der Ziel [63] 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

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

converting to a Lorentzian 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

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

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

$$S_\epsilon(\omega) = \int_0^\infty S_\alpha(\omega) f(\alpha) \, d\alpha$$

(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

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

(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 $1/f$-type 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

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

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

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

is bounded according to

$$\frac{\sigma_f^2}{|\omega|^\gamma} \leq S_x(\omega) \leq \frac{\sigma_f^2}{|\omega|^\gamma}$$

for some $0 < \sigma_f^2 < \sigma_f^2 < \infty$, and has ripple such that for all integers $k$

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

As $\Delta$ 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, studies [64] [65] [66] 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 [43] [46] developed an alternative ARMA-based model for $1/f$-like behavior from an engineering perspective.

Sec. 3.3 Nearly-$1/f$ Processes

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

$$V(s) \propto \frac{1}{s^{1/2}}.$$
values of \( \gamma \), pole-zero cancellation takes place. In particular, as \( \gamma \to 2 \), 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 \( \gamma \to 0 \), 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\( \omega \) dB/octave.

However, a 1/f synthesis filter must fall off at

\[
10 \log \omega^2 \approx 3\gamma \quad \text{dB/octave}
\]

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 3\( \gamma \) 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 by Pellegrini et al. [68], while details of a discrete-time implementation of Keshner's model appears in Corsini and Saletti [69]. A comparison of the two approaches is presented in Saletti [67]. In virtually all the simulations we present, the Corsini-Saletti implementation of Keshner's model is used to synthesize 1/f processes. In particular, the 1/f sample paths of Fig. 3.1 are obtained via this algorithm.

### 3.3.2 Wavelet-Based Models

In this section, we explore the relationship between orthonormal wavelet bases and nearly-1/f models. In particular, we 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-Loève-type expansions for 1/f-type processes [59][60]. That is, wavelet basis expansions in terms of uncorrelated random variables constitute very good models for 1/f-type behavior.

2Noise that falls off at 3dB/octave, which corresponds to \( \gamma = 1 \), is often referred to as "pink" noise. It arises in a variety of applications. For example, in professional audio systems, room equalization is often performed with a pink noise source.

Synthesis

In this section, we demonstrate that nearly-1/f behavior may be generated from orthonormal wavelet basis expansions in terms of a collections of uncorrelated wavelet coefficients. In particular, we establish the following theorem, an earlier version of which appears in Wornell [70], 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

\[
x(t) = \sum_{m} x^m \psi^m(t),
\]

where the \( x^m \) are a collection of mutually uncorrelated, zero-mean random variables with variances

\[
\text{var } x^m = \sigma^2 2^{-m}
\]

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

\[
S_x(\omega) = \sigma^2 \sum_{m} 2^{-m}|\Psi(2^{-m}\omega)|^2
\]

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

\[
\frac{\sigma^2}{|\omega|^\gamma} \leq S_x(\omega) \leq \frac{\sigma_0^2}{|\omega|^\gamma}
\]

for some 0 < \( \sigma_0^2 < \sigma^2 < \infty \), and has octave-spaced ripple, i.e., for any integer \( k \)

\[
|\omega|^\gamma S_x(\omega) = |2^k \omega|^\gamma S_x(2^k \omega).
\]

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 \( \sigma_0^2/\sigma_1^2 = 1.103 \).

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 2.3.1. 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

\[
\sigma_0^2 = \sigma^2 \pi^\gamma
\]

\[
\sigma_1^2 = \sigma^2 (2\pi)^\gamma.
\]

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

\[
\text{var } x^m = \sigma^2 2^{-m}
\]
corresponds to distributing power equally among the detail signals at all resolution scales, since we have for any \( m \)

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

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 \( \omega_c < |\omega| \leq \omega_0 \). 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 3.3.1, 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 \( \hat{x}(t) \) constructed in this way is no longer ergodic. Furthermore, if the coefficients \( x_m^* \) in Theorem 3.4 are chosen to be Gaussian, \( x(t) \) will be necessarily a Gaussian process, but \( \hat{x}(t) \) 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 multisresolution 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 \( \gamma > 2 \) requires higher regularity \((R > 1)\) is required. This can be verified experimentally as well. We find, for instance, that when we attempt to synthesize 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 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, we do not discuss this extension to the theorem primarily because there would appear to be relatively few, if any, physical processes of interest corresponding to negative \( \gamma \).
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 with Karhunen-Loève-type properties.

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 theorem, whose proof is outlined in Appendix B.4.

Theorem 3.5 Let $x(t)$ be a 1/f process whose spectral parameters, in the sense of Theorem 3.2, are $\sigma^2$ 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

$$E \left[ x_n^m x_{n'}^{m'} \right] = \frac{2^{-(m+m')/2}}{2\pi} \int_{-\infty}^{\infty} \frac{\sigma^2}{|\omega|^\gamma} |\Psi(\omega)|^2 e^{-j(2^{-m-n'} - 2^{-m}) \omega} d\omega$$

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

The principal shortcoming of this theorem is that it fails to establish conditions on the wavelet basis and $\gamma$ under which (3.40) is defined. Nevertheless, we may generally use Theorem 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 theorem is that we can show the variance of the $x_n^m$ to be of the form

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

where

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

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

Defining

$$\rho_{n,n'}^{m,m'} \triangleq \frac{E \left[ x_n^m x_{n'}^{m'} \right]}{\sqrt{\text{var} x_n^m \text{var} x_n^{m'}}}$$

as the normalized wavelet correlation, a second consequence is that the wavelet coefficients 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

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

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 2.3 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 $n = \ldots, -1, 0, 1, \ldots$, 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

$$2^{-m}n = 2^{-m'}n'.$$  

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 may readily establish that whenever (3.43) holds,

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

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.,

$$\psi^{(r)}(t) = 0, \quad r = 0, 1, \ldots, R - 1$$

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 to\(^3\)
\[
|\rho_{m,n}^{m',n'}| \sim \mathcal{O} \left( 2^{-m n} - 2^{-m' n'} \right)^{-\gamma/2} \tag{3.46}
\]
as
\[
|2^{-m n} - 2^{-m' n'}| \to \infty.
\]

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 leads 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 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

\[
\rho_{m,n+l}^{m,n} = \frac{\sigma^2}{\pi \sigma^2} \int_\pi 2^{\gamma \cos(\omega l)} d\omega \tag{3.47}
\]

where
\[
\sigma^2 = \begin{cases} 
\frac{(2^{1-\gamma} - 1)/(\pi^2(1 - \gamma))}{\gamma \neq 1} \\
\frac{(\ln 2)/(\pi)}{\gamma = 1}
\end{cases} \tag{3.48}
\]

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

\[
|\rho_{m,n+l}^{m,n}| \leq \frac{\sigma^2}{\pi \sigma^2} \frac{\gamma}{2^{\gamma + 1}} \left[ 1 + \frac{1}{2^{1+\gamma}} + \frac{1 + \gamma}{l \pi} \left[ 1 - \frac{1}{2^{2+\gamma}} \right] \right] \tag{3.49}
\]
valid for \(\gamma \geq 0\) and integer-valued \(l \geq 1\).

\(^3\)The ceiling function \(\lceil x \rceil\) denotes the smallest integer greater than or equal to \(x\).
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 the analysis result for the bandlimited basis may, in fact, be more broadly applicable.

Before concluding this analysis section, it is appropriate to point out 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 [44]; the interscale stationarity property was described by Flandrin [71] (after Vergassola and Frisch [72]). 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 [73]. We also mention that Flandrin [71] 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 related results that have emerged concerning the effectiveness of wavelet decompositions in decorrelating a broad class of smooth covariance kernels [74].

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 example involves economic data, and is depicted in Fig. 3.5. The second example involves physiological data, and is depicted in Fig. 3.6.

Focusing first on the economic data, Fig. 3.5 shows the time series corresponding to raw weekly Dow Jones Industrial Average data accumulated over the past approximately 80 years. As shown in Fig. 3.7(a), the sample-variance of wavelet coefficients from scale to scale obeys a geometric progression consistent with a $1/f$ process for which $\gamma \approx 2$. In Fig. 3.7(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 intercoefficient 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.

Turning our attention next to the physiological data, Fig. 3.6 shows a record of heart beat interarrival times for a healthy human patient corre-
Figure 3.7. Wavelet-based analysis of weekly Dow Jones Industrial Average data. The time-series is analyzed using a 5th-order Daubechies wavelet basis. (a) Scale-to-scale wavelet coefficient sample-variance progression. (b) Average magnitude of the normalized along-scale sample-correlation between wavelet coefficients.

Corresponding to approximately 11 hours of continuously acquired data. The quantization levels of the interarrival times are spaced 4 milliseconds apart. In this example, as shown in Fig. 3.8(a), the sample-variances of wavelet coefficients from scale to scale obey a geometric progression consistent with a $1/f$ process of $\gamma \approx 1$. 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 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 intercoefficient correlation in any wavelet-based analysis. Again, we caution that no attempt was made to study the correlation structure among coefficients between scales.

### 3.4 Summary

In this chapter, we focused on fractal random processes having the key property that their statistics are invariant to temporal dilations and contractions of the process, to within an amplitude factor. Of particular interest were the $1/f$ family of such statistically self-similar random processes, and we developed some important new models for $1/f$-type behavior in signals. After reviewing the traditional fractional Brownian motion model for $1/f$ behavior, we developed a powerful alternative frequency-based characterization for $1/f$ processes. Here we showed that, although they are generally nonstationary, $1/f$ processes have the special property that when bandpass filtered they always produce stationary outputs.

In the second half of the chapter we relaxed our model constraints, and considered nearly-$1/f$ processes. We began by reviewing traditional ARMA models for nearly-$1/f$ behavior, then turned our attention to developing wavelet-based representations for $1/f$-type processes. As our main result we demonstrated that orthonormal wavelet basis expansions are Karhunen-Loève-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 result, which has powerful implications, was supported both theoretically and empirically, and we presented examples involving both simulated and real data.

Collectively, our theoretical and empirical results suggest that the orthonormal wavelet transform is an extremely useful and convenient tool in the synthesis and analysis of $1/f$-type processes. In the next chapter we explore how the wavelet transform plays an equally valuable role in the processing of such signals.
4

Detection and Estimation with Fractal Processes

4.1 INTRODUCTION

Given the ubiquity of physical signals exhibiting 1/f-type behavior, there are many applications in which the need arises for 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 noise 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 noise. It is important to both recognize and accommodate such measurement noise in any algorithms for

\[ \text{Figure 3.8. Wavelet-based analysis of the heartbeat interarrival times for a healthy patient. The time-series is analyzed using a 5th-order Daubechies wavelet basis. (a) Scale-to-scale wavelet coefficient sample-variance progression. (b) Average magnitude of the normalized along-scale sample-correlation between wavelet coefficients.} \]
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 is signals that, when expanded into an orthonormal wavelet basis, yield coefficients that are effectively uncorrelated and obey the appropriate variance progression. That is, we 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 2.3, 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.

We routinely incorporate additive stationary white measurement noise to ensure the robustness of the algorithms we develop. Furthermore, most of these algorithms are 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 settings in which this assumption is physically quite reasonable. Furthermore, several of the algorithms we develop retain many of their important properties in the more general non-Gaussian case.

An important component of the algorithm development process is performance analysis, and this can take many forms. Accompanying each of the algorithms we develop is a set of basic though necessarily limited evaluations of its performance in various settings. These analyses both establish the essential viability of the algorithms and reveal some of their salient 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 [69]. 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, as a consequence, these simulations generally do not enable us to isolate the effects of modeling error alone.

frequencies below about 1 kHz, while at higher frequencies, it is white noise in the form of thermal (i.e., Johnson) and shot noise [75].

Sec. 4.2 1/f Synthesis and Whitening Filters

The wavelet-based implementation of each of the algorithms we develop requires that we select a suitable wavelet basis from among the large collection of candidates. However, given the apparent insensitivity of the wavelet-based model for $1/f$-type behavior to the choice of basis, for the 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 3.3.2 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, $0 < \gamma < 2$.

Before beginning, we note that a basic understanding of the fundamentals of estimation and detection theory is assumed of the reader in this chapter. A sufficiently comprehensive treatment for our purposes can be found in, e.g., Van Trees [76].

4.2 1/f SYNTHESIS AND WHITENING FILTERS

Many of the results on detection and estimation we 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 $\gamma$, 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

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

where, for notational convenience, we define

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

In turn, we may express the $x_n^m$ as

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

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

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

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 \( z_n^m \) to generate the output process \( x(t) \). This 1/f synthesis filter, defined via
\[
x(t) = W_d^{-1} \left\{ \sigma \beta^{-m/2} W_d \{ v(t) \} \right\},
\] (4.2)
is a linear filter whose kernel\(^2\) is
\[
k_n^m(t, \tau) = \sum_m \sum_n \psi_n^m(t) \sigma \beta^{-m/2} \psi_n^m(\tau).
\] (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 \( v(t) \) 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
\[
k_n^m(t, \tau) = \sum_m \sum_n \psi_n^m(t) \frac{1}{\sigma \beta^{-m/2}} \psi_n^m(\tau).
\] (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 white 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 3.3.2.

As discussed at the outset of the chapter, any 1/f-type process we consider are invariably accompanied by an additive stationary white observation noise component. Consequently, we 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-plus-white processes are characterized by the respective kernels
\[
\kappa_n(t, \tau) = \sum_m \sum_n \psi_n^m(t) \sigma_m \psi_n^m(\tau)
\] (4.4a)

where \( \sigma_m > 0 \) is defined by
\[
\sigma_m^2 = \sigma^2 \beta^{-m} + \sigma_v^2
\] (4.5)
and \( \sigma_v^2 \) is the spectral density of the white noise component. The canonical wavelet-based realization of these filters is depicted in Fig. 4.1.

### 4.3 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 [77]. Since we typically lack a priori knowledge of the spectral
density of the noise, we consider more specifically 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 using the relationships developed in Chapter 3. Robust estimation of the fractal dimension of 1/f processes is important in a number of applications requiring 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 [47] [58], fractal dimension can be useful in distinguishing among various man-made and natural objects.

While several approaches to the fractal dimension estimation problem can be found in the literature (see [58], [78], [79], and the references therein), a traditional approach with these approaches has been their inability 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 [58]. Since noise is inherently present in any real data, this lack of robustness has limited the usefulness of these algorithms. In this section we describe 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. 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 stress 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

\[
r(t) = x(t) + w(t), \quad -\infty < t < \infty.
\]

(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, i.e., via

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

However, in practice, these coefficients are more typically obtained by applying the computationally efficient DWT to the samples of a segment of data that is both time-limited and resolution-limited, as described in Section 2.3.3. Let us assume that the finite set of available distinct scales, \( M \), is, in increasing order,

\[
M = \{m_1, m_2, \ldots, m_M\},
\]

(4.7a)

and that at each scale \( m \) the set of available coefficients \( N(m) \) is

\[
N(m) = \{n_1(m), n_2(m), \ldots, n_{N(m)}(m)\}.
\]

(4.7b)

Hence, the data available to the estimation algorithm are

\[
r = \{r_n^m \in \mathcal{R} \} = \{r_n^m, m \in M, n \in N(m)\}.
\]

(4.8)

We remark before proceeding that, based on the discussion in Section 2.3.4, for an implementation via the DWT with \( N = N_02^M \) samples of observed data, we have, typically,

\[
M = \{1, 2, \ldots, M\},
\]

(4.9a)

\[
N(m) = \{1, 2, \ldots, N_02^{m-1}\},
\]

(4.9b)

where \( N_0 \) is a constant that depends on the length of the filter \( h[n] \). While many of the results we derive are applicable to the more general scenario, we 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

\[
\text{var} r_n^m = \sigma_r^2 = \sigma_x^23^{-m} + \sigma_w^2
\]

where \( \beta \) is defined in terms of the spectral exponent \( \gamma \) of the 1/f process according to (4.1). Hence, it is the parameter set

\[
\Theta = (\beta, \sigma_x^2, \sigma_w^2)
\]

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

\[
\gamma_{\text{ML}} = \log_2 \beta_{\text{ML}}
\]

(4.10a)

\[
D_{\text{ML}} = (5 - \gamma_{\text{ML}})/2
\]

(4.10b)

\[
H_{\text{ML}} = (\gamma_{\text{ML}} - 1)/2.
\]

(4.10c)

\footnote{Note that without loss of generality we may assume \( N(m) \neq \emptyset \) for any \( m \), or else the corresponding scale \( m \) could be deleted from \( M \).}
Proceeding, we express the likelihood as a function of the parameters by

$$L(\Theta) = p_\theta(T; \Theta) = \prod_{m \in \mathcal{M}} \frac{1}{\sqrt{2\pi \sigma_m^2}} \exp \left( -\frac{(r_n^m)^2}{2\sigma_m^2} \right)$$

for which the log-likelihood function is

$$L(\Theta) = \log p_\theta(T; \Theta) = -\frac{1}{2} \sum_{m \in \mathcal{M}} \left( \frac{1}{\sigma_m^2} (r_n^m)^2 + \ln(2\pi \sigma_m^2) \right).$$

Equivalently,

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

where the M sample-variances

$$\sigma_m^2 = \frac{1}{N(m)} \sum_{n \in \mathcal{N}(m)} (r_n^m)^2$$

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 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 \(\sigma_w^2\) 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

$$\tilde{\beta} = \beta$$

$$\tilde{\sigma} = \sigma^2 \beta^{m-1}$$

$$\tilde{\sigma_w} = \sigma_w^2$$

for which the observations correspond to positive scales

$$\mathcal{M} = \{1, m_1 - m_1 + 1, \ldots, m_M - m_1 + 1\}$$

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

### 4.3.1 Case I: \(\beta, \sigma^2, \sigma_w^2\) Unknown

Differentiating \(L(\Theta)\) with respect to \(\sigma_w^2, \sigma^2,\) and \(\beta\), respectively, it follows that the stationary points of \(L(\Theta)\) are given as the solutions to the equations

$$\sum_{m \in \mathcal{M}} T_m = 0$$

$$\sum_{m \in \mathcal{M}} m T_m = 0$$

$$\sum_{m \in \mathcal{M}} m^2 T_m = 0$$

where

$$T_m = \frac{N(m)}{\sigma_m^2} \left( 1 - \frac{\sigma_w^2}{\sigma_m^2} \right)$$

However, these equations are difficult to solve, except in special cases. Consequently, we utilize an iterative estimate-maximize (EM) algorithm [80].

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 \(\tilde{\beta}, \tilde{\sigma}^2,\) and \(\tilde{\sigma}_w^2\) generated on the \(lth\) iteration by \(\beta[l], \sigma^{2[l]},\) and \(\sigma_w^{2[l]}\), respectively.

**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 \(\beta[l], \sigma^{2[l]},\) and \(\sigma_w^{2[l]}\):

$$S_m^w(\Theta[l]) = A_m(\Theta[l]) + B_m(\Theta[l])\sigma_w^2$$

$$S_m^s(\Theta[l]) = A_m(\Theta[l]) + B_m(\Theta[l])\sigma_s^2$$

where

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

$$B_m(\Theta[l]) = \left( \frac{\sigma_w^2[l]}{\sigma_w^2[l] + \sigma^{2[l]}[\beta[l]]^{-m}} \right)^2$$

**M step:** This step reduces to using these estimates and noise variance estimates to obtain the new parameter estimates \(\tilde{\beta}^{[l]+}, \tilde{\sigma}^{2[l]+},\) and \(\tilde{\sigma}_w^{2[l]+}\):

$$\beta^{[l]+} = \sum_{m \in \mathcal{M}} C_n N(m) S_m^w(\Theta[l])^m = 0$$

$$\sigma^{2[l]+} = \sum_{m \in \mathcal{M}} N(m) S_m^s(\Theta[l])^m$$

$$\sigma_w^{2[l]+} = \sum_{m \in \mathcal{M}} N(m)$$

Sec. 4.3 Parameter Estimation for 1/f Signals
where

\[ C_m \triangleq \frac{1}{\sum_{m \in M} mN(m)} \sum_{m \in M} mN(m). \]  

**(4.16)**

### 4.3.2 Case II: \( \beta, \sigma^2 \) Unknown; \( \sigma^2_w \) Known

If \( \sigma^2_w \) is known, the above algorithm simplifies somewhat. In particular, we may omit the estimation (4.15c) and replace occurrences of \( \hat{\sigma}_m^2(\Theta) \) in the algorithm with the true value \( \sigma^2_w \). This eliminates the need to compute \( S^w_m(\Theta) \) and, hence, \( B^w_m(\Theta) \). The resulting algorithm is as follows.

**E step:** Estimate the signal portion of the wavelet coefficient variances at each scale \( m \in M \) using current estimates of the parameters \( \beta \) and \( \sigma^2_w \):

\[
S^s_m(\Theta^l) = A^s_m(\Theta^l) + B^s_m(\Theta^l)\sigma^2_w
\]

where

\[
A^s_m(\Theta^l) = \frac{\sigma^2_w \cdot \sigma^2_w | \beta_n |^{-m}}{\sigma^2_w + \sigma^2_w | \beta_n |^{-m}}
\]

\[
B^s_m(\Theta^l) = \left( \frac{\sigma^2_w | \beta_n |^{-m}}{\sigma^2_w + \sigma^2_w | \beta_n |^{-m}} \right)^2.
\]

**M step:** Use these signal variance estimates to obtain the new parameter estimates \( \beta^{l+1} \) and \( \sigma^2_w^{l+1} \):

\[
\beta^{l+1} = \sum_{m \in M} \frac{C_m N(m)S^s_m(\Theta^l)\beta^m}{\sum_{m \in M} N(m)} = 0
\]

\[
\sigma^2_w^{l+1} = \frac{\sum_{m \in M} N(m)S^s_m(\Theta^l) | \beta^{l+1} |^{-m}}{\sum_{m \in M} N(m)}
\]

where \( C_m \) is as in (4.16).

### 4.3.3 Case III: \( \beta, \sigma^2 \) Unknown; \( \sigma^2_w = 0 \)

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

\[
\hat{\beta}_ML \left| \sigma^2_w = 0 \right. = \sum_{m \in M} C_m N(m)\hat{\sigma}_m^2|\beta|^{-m} = 0
\]

\[
\sigma^2_w = 0
\]

### 4.3.4 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:

\[
\beta = \hat{\beta}_ML |_{\sigma^2_w = 0}
\]

\[
\sigma^2 = \hat{\sigma}_ML^2 |_{\sigma^2_w = 0}
\]

\[
\sigma^2_w = 0
\]

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 Appendix C. We begin by differentiating the likelihood function to find equations for its stationary points. This leads to a pair of equations in terms of \( \sigma^2 \) and \( \beta \). Eliminating \( \sigma^2 \) from these equations is straightforward and gives (4.20a) directly. Having determined \( \beta_{ML} \) as the solution to this polynomial equation, \( \sigma^2_{ML} \) 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 \( \beta_{ML} \). 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 \( \beta_{ML} \). For instance, when we expect \( 0 < \beta < 2 \), an appropriate initial interval is \( 1 < \beta < 4 \). Naturally, with some caution, Newton iterations \([U]\) 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.
That they are saddle points is rather fortunate, for the only way they are reached is if the starting value for any one of \( \beta, \sigma^2, \sigma_m^2 \) is chosen to be exactly zero. Given arbitrarily small positive choices for these initial parameters, the algorithm iterates 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 naturally affects the rate of convergence of the algorithm. Moreover, it should be noted that the EM algorithm converges substantially faster for the case in which \( \alpha \) is known. In essence, for the general algorithm much of the iteration is spent locating the noise threshold in the data.

Turning now 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 [82]. It is also the case that at least in some higher signal-to-noise ratio (SNR) scenarios, the Cramer-Rao bounds closely approximate the true estimation error variances.

To compute the Cramér-Rao bounds for the estimates of \( \gamma, \sigma^2, \) and \( \sigma_m^2 \), we construct the corresponding Fisher matrix

\[
I = \sum_{m \in M} \frac{N(m)}{2(\sigma_m^2)^2} \begin{bmatrix}
[\ln 2 \sigma^2 \beta^2 - \ln 2(\sigma^2)^2] & -\ln 2(\sigma^2)^2 \\
-\ln 2(\sigma^2)^2 & [\ln 2(\sigma^2)^2]^{1/2} \beta_m^{-1} - 1
\end{bmatrix}
\]

from which we get

\[
\begin{align*}
\text{var} \, \hat{\gamma} & \geq I^{11} \\
\text{var} \, \hat{\sigma}^2 & \geq I^{22} \\
\text{var} \, \hat{\sigma}_m^2 & \geq I^{33}
\end{align*}
\]

for any unbiased estimates \( \hat{\gamma}, \hat{\sigma}^2, \hat{\sigma}_m^2 \), and where \( I^{ik} \) is the \( k \)-th element on the diagonal of \( I^{-1} \). 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 \( \sigma_m^2 \) is known, the Fisher information matrix simplifies to the upper submatrix

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

for any unbiased estimates \( \gamma, \sigma^2, \sigma_m^2 \), and where \( I^{ik} \) is the \( k \)-th element on the diagonal of \( I^{-1} \). 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 \( \sigma_m^2 \) is known, the Fisher information matrix simplifies to the upper submatrix

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

from which we get

\[
\begin{align*}
\text{var} \, \hat{\gamma} & \geq I^{11} \\
\text{var} \, \hat{\sigma}^2 & \geq I^{22} \\
\text{var} \, \hat{\sigma}_m^2 & \geq I^{33}
\end{align*}
\]

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 \( \sigma_m^2 = 0 \), we get still further simplification, and we can write

\[
I = \begin{bmatrix}
(\ln 2)^2/2 \sum_{m \in M} m^2 N(m) & -(\ln 2)/2(\sigma^2) \sum_{m \in M} m N(m) \\
-(\ln 2)/2(\sigma^2) \sum_{m \in M} m N(m) & 1/2(\sigma^2) \sum_{m \in M} N(m)
\end{bmatrix}
\]

from which we get

\[
\begin{align*}
\text{var} \, \hat{\gamma} & \geq 2/[(\ln 2)^2/2] \sum_{m \in M} N(m) \\
\text{var} \, \hat{\sigma}^2 & \geq 2/2 \sum_{m \in M} m^2 N(m)
\end{align*}
\]

where

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

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

\[
\begin{align*}
\text{var} \, \hat{\gamma}_{ML} & \sim 2/[(\ln 2)^2 N] \\
\text{var} \, \hat{\sigma}^2_{ML} & \sim 2(\log_2 N)^2/N
\end{align*}
\]

where \( N \) is the number of observation samples.

### 4.3.5 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 observation sequence as well as the spectral exponent \( \gamma \) 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, the RMS error in the estimates of $\gamma$ and $\sigma^2$ is plotted for various values of $\gamma$ 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 $\gamma$ and are achieved asymptotically. For increasing values of $\gamma$, the results suggest that the bounds are attained at increasing SNR thresholds.

In Fig. 4.3, the RMS error in the estimates of $\gamma$ and $\sigma^2$ is plotted for various values of $\gamma$ 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 $\sigma^2$, that the convergence toward the true parameters can be rather slow. Note, too, that a rather large amount of data is required before the relative estimation error in $\sigma^2$ 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 $\gamma$ 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 $\gamma$ 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. It is useful to point out that, as is the case with most tracking algorithms, using a wider estimation window would reduce the variance in the parameter estimates with each half of the waveform, but at the expense of an increase in the width of the transition zone.

### 4.4 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 [27]. 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
Detection and Estimation with Fractal Processes  Chap. 4

\[ SNR = 20 \text{ dB} \]

\[ y = 0.33 \]
\[ y = 1.00 \]
\[ y = 1.67 \]

\[ I \]

\[ {\text{Data Length } N} \]

\[ A \]

\[ r = 1.00 \]

\[ y = 1.67 \]

\[ q \]

\[ :: \]

\[ A \]

\[ B \]

\[ -3 \]

\[ \text{Fig. 4.3. RMS Errors in the estimates of the signal parameters as a function of the data length } N \text{ of the observations. Again, the symbols associated with each } y \text{ mark the actual empirical measurements; dashed lines are provided as visual aides only. (a) Absolute RMS error in } \hat{\gamma} \text{. (b) Percentage RMS error in } \hat{\alpha}. \]

\[ \text{Sec. 4.4 Smoothing of } 1/f \text{ Signals} \]

\[ \text{window size } = 16 \text{ 384} \]

\[ \text{SNR } = \infty \]

\[ \text{Fig. 4.4. Tracking the time-varying spectral exponent } \gamma \text{ of a noise-free } 1/f \text{-type signal. For the left half of the signal, } \gamma = 0.90, \text{ while for right half, } \gamma = 1.10. \text{ (a) } 1/f \text{ signal with step change in } \gamma. \text{ (b) Estimate of } \gamma \text{ under a time-limited window.} \]

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 is capable of yielding signal estimates with a smaller mean-square error [76].

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 assume in our derivation that the signal and noise parameters $\beta, \sigma^2, \sigma^2_w$ are all known, though, in practice they are typically estimated using the parameter estimation 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 $\sigma^2_w$ 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

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

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

$$E[x_n^m | r_n^m] = \frac{\sigma^2\beta^m}{\sigma^2\beta^m + \sigma^2_w} r_n^m.$$  \hspace{1cm} (4.25)

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

$$\hat{x}(t) = \sum_{m, n} \hat{x}_n^m \psi_n^m(t) = \sum_{m, n \in \mathbb{R}} \left[ \frac{\sigma^2\beta^m}{\sigma^2\beta^m + \sigma^2_w} \right] r_n^m \psi_n^m(t).$$  \hspace{1cm} (4.26)

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

$$\frac{\sigma^2\beta^m}{\sigma^2\beta^m + \sigma^2_w}$$

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 (4.14), 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 4.2 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 $k_w(t, \tau)$ given by (4.4b) to generate an intermediate white innovations process $v(t)$ whose wavelet coefficients are

$$v_n^m = \frac{r_n^m}{\sigma_m}.$$  \hspace{1cm} (4.27)

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

$$k_i(t, \tau) = \sum_m \sum_n \psi_n^m(t) \left[ \frac{\sigma^2\beta^m}{\sigma^2\beta^m + \sigma^2_w} \right] \psi_n^m(\tau),$$

which is stationary and white.

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

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

is infinite. Nevertheless, when we define

$$x_R(t) = \sum_{m, n \in \mathbb{R}} x_n^m \psi_n^m(t)$$

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 $x_R(t)$ as

$$\varepsilon = \int_{-\infty}^{\infty} E[(\hat{x}(t) - x_R(t))^2] \, dt = \sum_{m, n \in \mathbb{R}} E[(\hat{x}_n^m - x_n^m)^2] + \sum_{m, n \in \mathbb{R}} E[\text{var}(x_n^m \psi_n^m)]$$

which, through routine manipulation, reduces to

$$\varepsilon = \sum_{m \in \mathbb{M}} N(m) \left[ \frac{\sigma^2\beta^m}{\sigma^2\beta^m + \sigma^2_w} \right].$$  \hspace{1cm} (4.28)
As a final comment, note that while we do not develop the 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 dominates, while at low frequencies the \(1/f\) signal spectrum dominates. In fact, at sufficiently low frequencies, there is always 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. Moreover, this low-pass filtering is effectively implemented on a logarithmic frequency scale—the scale which is most natural for these processes.

### 4.4.1 Simulations

For the simulations of this section, discrete samples of resolution-limited Gaussian \(1/f\) processes embedded in Gaussian white noise are synthesized. In general, the SNR of the observations sequence is varied as well as the spectral exponent \(\gamma\) of the underlying \(1/f\) processes. Parameter estimation is then performed, followed by signal estimation, using algorithms for the most general scenario, corresponding to the case in which all signal and noise parameters \(\beta, \sigma^2, \sigma_n^2\) are unknown. Note that by using the estimated parameters in the signal estimation algorithm, these 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. This limitation is not serious, however, since the quality of the signal estimation is generally rather insensitive to errors in the parameter estimates used.

In Fig. 4.6, the SNR gain of the smoothed signal estimates is plotted for various values of \(\gamma\) 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 \(\gamma\) for a given SNR. This is to be expected since for larger values of \(\gamma\) there is more signal energy at coarser scales and correspondingly less at finer scales where the noise predominates and the most attenuation takes place.

We conclude this section with a signal estimation example. Fig. 4.7 shows 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 is 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.

### 4.5 COHERENT DETECTION IN \(1/f\) NOISE

In this section the problem of detecting a known signal of finite energy in a background of Gaussian \(1/f\) and white noise is considered. The detection algorithms we develop may be used in a variety of applications; they may be exploited, for example, in communication and pattern recognition systems.
In our approach, we explicitly include stationary white Gaussian measurement noise in our model. In general, this refinement improves the robustness properties of the resulting algorithms and, in particular, precludes certain singular detection scenarios. As will become apparent, the wavelet-based approach we develop 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 [76]. Specifically, 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 $H_0$ we observe only the background noise, i.e.,

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

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 assume that $w(t)$ and $x(t)$ are statistically independent processes under either hypothesis, and also that our observations generally extend over the infinite interval $-\infty < t < \infty$. We then seek 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 [76].

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

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

as

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

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 $\sigma_w^2$ and $\sigma_x^2$, respectively.

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

$$p_{H_1}(r) = \prod_{m,n} \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp \left[ -\frac{(r_n^m - \sqrt{E_0} s_n^m)^2}{2\sigma_n^2} \right]$$
$$p_{H_0}(r) = \prod_{m,n} \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp \left[ -\frac{(r_n^m)^2}{2\sigma_n^2} \right],$$

the likelihood ratio

$$\frac{p_{H_1}(r)}{p_{H_0}(r)}$$
Detection and Estimation with Fractal Processes  Chap. 4

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.

can be simplified substantially to yield a test of the form

\[
\ell = \sum_n \sum_m r_m^* s_m^* \frac{H_1}{H_0} \gtrless \alpha
\]

(4.29)

where \( \alpha \) 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 \( \ell \) 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 then correlating the respective outputs \( r_n(t) \) and \( s_n(t) \). It is straightforward to verify this implementation: since the prewhitened signals \( r_n(t) \) and \( s_n(t) \) have wavelet coefficients \( r_n^\sigma_m \) and \( \sqrt{E_0}s_n^\sigma_m \), respectively, it suffices to recognize the expression for \( \ell \) in (4.29) as the inner product between \( s_n(t)/\sqrt{E_0} \) and \( r_n(t) \), which allows us to rewrite (4.29) as

\[
\ell = \int_{-\infty}^{\infty} r_n(t) s_n(t)/\sqrt{E_0} \ dt \gtrless \alpha.
\]

This is, of course, a canonical form receiver for optimal detection in the presence of colored noise as described in [76].

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

\[
d^2 \triangleq \int_{-\infty}^{\infty} s_n^2(t) \ dt = E_0 \sum_n \frac{(s_n^m)^2}{\sigma_n^2} \]

(4.30)

then

\[
E[\ell|H_0] = 0, \quad E[\ell|H_1] = d^2/\sqrt{E_0}, \quad \text{var}(\ell|H_0) = \text{var}(\ell|H_1) = d^2/E_0.
\]

Hence, expressing our arbitrary threshold in the form

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

for some \( 0 < \eta < \infty \), the performance of the test can be described in terms of the detection and false alarm probabilities, respectively

\[
P_D = Q \left( \frac{\ln \eta - \frac{d}{2}}{\frac{d}{2}} \right) \]

(4.31a)

\[
P_F = Q \left( \frac{\ln \eta + \frac{d}{2}}{\frac{d}{2}} \right)
\]

(4.31b)

where

\[
Q(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-v^2/2} dv.
\]

(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 \( d \).
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 attempts 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. An example of how this can be accommodated in a communications scenario is described in Wornell [83].

### 4.6 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 4.3. Indeed, parameter estimation can be viewed as distinguishing among an arbitrarily large number of 1/f processes with incrementally different parameters. Nevertheless, as we will 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

\[
H_{0} : r(t) = \hat{x}(t) + w(t) \quad (4.33a)
\]

\[
H_{1} : r(t) = \tilde{x}(t) + w(t) \quad (4.33b)
\]

where \( \hat{x}(t) \) and \( \tilde{x}(t) \) are Gaussian 1/f processes\(^{4}\) with distinct parameters and \( w(t) \) is a white measurement noise, statistically independent of \( \hat{x}(t) \) or \( \tilde{x}(t) \), whose variance is the same under both hypotheses. For this test we develop a minimum probability of error \([\Pr(\varepsilon)]\) decision rule under the assumption of equally likely hypotheses.

\(^{4}\)In this section, the notations ' and " are used 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 \( " \).

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

\[
H_{0} : r_{n}^{m} = \hat{x}_{n}^{m} + w_{n}^{m}
\]

\[
H_{1} : r_{n}^{m} = \tilde{x}_{n}^{m} + w_{n}^{m}
\]

we model the \( r_{n}^{m} \) under each hypothesis as a collection of zero-mean statistically independent Gaussian random variables with variances

\[
\text{var} \{r_{n}^{m}|H_{0}\} = \sigma_{m}^{2} = \sigma_{m}^{2} \hat{\beta}^{-m} + \sigma_{w}^{2}
\]

\[
\text{var} \{r_{n}^{m}|H_{1}\} = \tilde{\sigma}_{m}^{2} = \sigma_{m}^{2} \tilde{\beta}^{-m} + \sigma_{w}^{2}
\]

where

\[
\hat{\beta} = 2^{b}
\]

\[
\tilde{\beta} = 2^{c}
\]

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

\[
r = \{r_{m}^{n} \in \mathcal{R}\} = \{r_{m}^{n}, m \in \mathcal{M}, n \in \mathcal{N}(m)\},
\]

where \( \mathcal{M} \) and \( \mathcal{N}(m) \) are as defined in (4.7), are available. In fact, as we will 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 assume the 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 2.3.4.

The likelihood ratio test for this problem can, through straightforward manipulation, be simplified to a test of the form

\[
\ell = \frac{1}{2} \sum_{m \in \mathcal{M}} \sum_{n \in \mathcal{N}(m)} \left\{ \frac{1}{\hat{\sigma}_{m}^{2}} - \frac{1}{\tilde{\sigma}_{m}^{2}} \right\} \left( \frac{r_{n}^{m}}{\hat{\sigma}_{m}^{2}} \right)^{2} - \ln \left( \frac{\hat{\sigma}_{m}^{2}}{\tilde{\sigma}_{m}^{2}} \right) \geq 0.
\]

It is straightforward to show that this test can be implemented in the canonical form shown in Fig. 4.10. Here the observations \( r(t) \) are processed by 1/f-plus-white whitening filters corresponding to each hypothesis, for which the respective kernels are

\[
\kappa_{w}(t, \tau) = \sum_{m} \sum_{n} \psi_{m}^{n}(t) \frac{1}{\hat{\sigma}_{m}} \psi_{n}^{m}(\tau)
\]

\[
\tilde{\kappa}_{w}(t, \tau) = \sum_{m} \sum_{n} \psi_{m}^{n}(t) \frac{1}{\tilde{\sigma}_{m}} \psi_{n}^{m}(\tau)
\]

Consequently, only one of the residual processes \( \tilde{\psi}(t) \) and \( \hat{\psi}(t) \) is white, depending on which hypothesis is true. To decide between the two hypotheses,
Detection and Estimation with Fractal Processes  Chap. 4

Figure 4.10. A canonical form implementation of the optimal receiver for discriminating between 1/f models with distinct parameters based on noisy observations \( r(t) \).

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 \( \ell \) 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 Van Trees [76] based upon the use of the Chernoff bound. Specifically, defining

\[
\mu(s) = \ln E \left[ e^{s \ell} | H_0 \right],
\]

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

\[
\Pr(\varepsilon) \leq \frac{1}{2} e^{s* \mu(s*)},
\]

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 \( \ell \) as Gaussian via a central limit theorem (CLT) argument, we can also obtain the following asymptotic expression for the error probability

\[
\Pr(\varepsilon) \approx \frac{1}{2s*(1 - s*)} \sqrt{2\pi \mu''(s*)},
\]

which is a more optimistic and accurate estimate of the achievable performance [76].

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

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

\[
\mu'(s) = \frac{1}{2} \sum_{m \in M} N(m) \left\{ \ln \frac{\hat{\sigma}_m^2}{\sigma_m^2} - \ln \left( \frac{\hat{\sigma}_m^2}{\sigma_m^2} + (1 - s) \right) \right\} \cdot (4.38b)
\]

\[
\mu''(s) = \frac{1}{2} \sum_{m \in M} N(m) \left\{ \frac{\hat{\sigma}_m^2}{\sigma_m^2} - \frac{1}{\frac{\sigma_m^2}{\sigma_m^2} + (1 - s)} \right\}^2. \quad (4.38c)
\]

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 \( 0 \leq s \leq 1 \).

4.6.1 Simulations

In this section, we obtain, via (4.36) and (4.37), numerical estimates of the probability of error performance. As our scenario, we assume that coefficients \( r_n \) 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 the simulations, we consider the ability of the optimal receiver to distinguish between 1/f processes of different spectral exponents \( \gamma \) (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 \( \sigma^2 \) such that the variance of the observations is identical under either hypothesis.

In the 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 \( \gamma \)-dependent SNR,
Pr(ε) 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 ω = 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 a 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.

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 Δγ = 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, observe that the best performance is achieved for the smallest values of γ. In subsequent tests, we restrict our attention to performance in this high SNR regime.

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 one 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, better performance is achieved for smaller 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 Pr(ε) scenarios, (4.36) constitutes a more useful measure of system performance than does (4.37). This behavior is illustrated in Figs. 4.14, 4.15, and 4.16 for hypotheses in the γ = 1 regime. Note that only at sufficiently high SNR, data lengths, and parameter separations does the CLT-based approximation actually yield a Pr(ε) estimate that is below the bound. From these plots 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 → ∞. Nevertheless, the fact that the rate of change of Pr(ε) with respect to SNR, data length N, and parameter separation Δγ has a similar form for both the bound and the approximation suggests that the Chernoff bound provides a meaningful estimate of achievable performance.

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
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 $\Delta$ symbols correspond to actual estimates; the lines are provided as visual aids 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).
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 $\hat{c}_t^2$, $\hat{c}_u^2$ and $\sigma_u^2$ would be unwanted parameters of the problem. In this case, a solution could be obtained using a generalized likelihood ratio test [76] of the form

$$
\max_{c_t^2, c_u^2} \prod_{m,n \in \mathbb{R}} \frac{1}{\sqrt{2\pi \sigma_m^2}} \exp \left\{ \frac{(r_{m,n}^u)^2}{2\sigma_m^2} \right\} H_1 \\
\max_{c_t^2, c_u^2} \prod_{m,n \in \mathbb{R}} \frac{1}{\sqrt{2\pi \sigma_m^2}} \exp \left\{ \frac{(r_{m,n}^u)^2}{2\sigma_m^2} \right\} H_0
$$

(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 4.3. 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.

### 4.7 ALTERNATIVE APPROACHES AND RELATED DEVELOPMENTS

Until relatively recently, 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. This section contains a summary of at least some of these results, though the list here is certainly not comprehensive.

One example is the work described in Barton and Poor [57], which considers problems of detection in the presence of fractional Gaussian noise using reproducing kernel Hilbert space theory. Using this framework, both infinite- and finite-interval whitening filters are developed for this class of 1/f noises, which, in turn, yields 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. A broad overview of work in this area is contained in Basseville et al. [84] and the references therein. In that work, the authors develop a tree/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,” leads 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 developed using this framework are directly applicable to 1/f processes as shown in, e.g., Chou [85].

There has also been progress in exploiting a discrete-time analog of fractional Brownian motion to obtain useful parameter and signal estimation algorithms. Results in this area are described in, e.g., Deriche and Tewfik [86] [87]. In addition, possible refinements to the estimation algorithms developed in this chapter are described in Kaplan and Kuo [88]. Finally, some extensions and generalizations of the algorithms in this chapter are developed in Lam [89].

### 4.8 SUMMARY

In this chapter, we exploited the efficiency of wavelet basis expansions for the 1/f family of fractal random processes to develop solutions to some fundamental problems of optimal detection and estimation involving 1/f-type signals. As a foundation, we first derived wavelet-based synthesis and whitening filters for 1/f processes that formed the basis for essentially all the algorithms derived in the chapter.

We then proceeded to develop maximum likelihood algorithms for estimating the parameters of 1/f processes from noise-corrupted observations given various degrees of a priori knowledge. These included algorithms for robust fractal dimension estimation, which are useful in a wide range of applications. Next, we developed complementary minimum mean-square error signal smoothing algorithms for separating and extracting 1/f processes from noisy observations. By exploiting our wavelet-based framework, we obtained algorithms that are extremely computationally efficient and highly practicable.

In many applications, it is the noise that has the 1/f characteristics rather than the signal of interest. Such is the case, for example, in a variety of communications applications where information-bearing waveforms are transmitted in a combination of 1/f and white backgrounds. Motivated by these kinds of scenarios, we developed efficient wavelet-based algorithms for optimally detecting known signals in such backgrounds.

Finally, we addressed a common signal classification problem involving 1/f processes—specifically, discriminating between 1/f signals with different parameters. In this case, too, efficient minimum probability of error decision strategies were developed via a wavelet domain formulation.
Throughout the chapter, a variety of properties and insightful interpretations of the resulting algorithms were developed, and their performance characteristics were explored both analytically and empirically.

Finally, it is important to appreciate that many other algorithms, such as those for distinguishing among superimposed $1/f$ signals and for detecting $1/f$ signals with unknown parameters, can be similarly derived using the methods developed in this chapter. In addition, straightforward generalizations of many of the algorithms to two- and higher-dimensional data such as imagery can also be developed.

5

Deterministically Self-Similar Signals

5.1 INTRODUCTION

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

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

(5.1)

for all $a > 0$, are generally referred to in mathematics as homogeneuous functions, in particular 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 [90], homogeneous functions can be parameterized with only a few constants. As such, they constitute a rather limited class of signal models for many engineering applications.

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

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

(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, and our treatment follows that in Wornell and Oppenheim [91]. We will typically use the generic term “homogeneous signal” to refer to signals satisfying (5.2). However, when there is risk of confusion in our subsequent development we will specifically refer to signals satisfying (5.2) as bihomogeneous.
Homogeneous signals constitute an interesting and potentially valuable class of signals for use in, for example, a variety of communications-based applications. As an illustration of potential, in Chapter 6 we explore their use in developing a diversity strategy for embedding information into a waveform “on all time scales.” As a consequence of their intrinsic self-similarity, these waveforms have the property that an arbitrarily short duration time-segment is sufficient to recover the entire waveform, and hence the embedded information, given adequate bandwidth. Likewise an arbitrarily low-bandwidth approximation to the waveform is sufficient to recover the undistorted waveform, and again the embedded information, given adequate duration. Furthermore, we will see that these homogeneous waveforms have fractal properties 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 diversity paradigm for communication over highly unreliable channels of uncertain duration, bandwidth, and SNR, as well as in a variety of other contexts. We explore these and other issues, including implementation, in the next chapter. In the meantime, we turn our attention to developing a convenient and efficient mathematical framework for characterizing homogeneous signals that we will exploit.

Some important classes of homogeneous signals have spectral characteristics very much like those of 1/f processes, and, in fact, have fractal properties as well. Specifically, while all nontrivial homogeneous signals have infinite energy and many have infinite power, we will see that there are in fact classes of these 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. These are the homogeneous signals of interest in this chapter. We distinguish between these two classes of such signals in our subsequent treatment, denoting them energy-dominated and power-dominated homogeneous signals, respectively.

We begin our theoretical development by formalizing our notion of an energy-dominated homogeneous signal, and constructing vector space characterizations. In turn, these lead to some powerful constructions of orthonormal “self-similar” bases for homogeneous signals. In the process, 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.

Before proceeding, we point out that our development relies heavily on a rather natural and efficient vector space perspective. In addition to facilitating the derivation of key results, this approach leads to powerful geometrical interpretations. Accessible treatments of the appropriate mathematical background can be found in, e.g., portions of Naylor and Sell [28] or Reed and Simon [29].

### 5.2 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 3.2. Specifically, we choose the following.

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

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

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

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

Some preliminary remarks regarding this definition are worthwhile at this point. First, we note that the choice of passband edges at $\pi$ and $2\pi$ in our definition is, in fact, 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.

It is also worth noting 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 are not always able to actually “plot” signals of this class, we are able to suitably characterize such functions in some useful ways. We 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

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

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

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

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

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

(5.6)
Since in this theorem $X(\omega)$ does not depend on $\omega_L$ or $\omega_L^\prime$, this function may be interpreted as the generalized Fourier transform of the infinite-energy signal $x(t)$. 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 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 because (5.5) excludes $\omega = 0$ and $\omega = \infty$, knowledge of $X(\omega)$ does not uniquely specify $x(t) \in E^H$; i.e., the mapping

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

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 behavior in our subsequent theoretical development, all signals having a common $X(\omega)$ are be combined into an equivalence class. For example, two homogeneous functions $f(t)$ and $g(t)$ are equivalent if they differ by a homogeneous function whose frequency content is concentrated at the origin, such as $lH$ in the case that $H$ is an integer.

Because the dyadic self-similarity property (5.2) of bihomogeneous signals is very similar to the dyadic scaling relationship between basis functions in an orthonormal wavelet basis, wavelets provide a particularly nice representation for this family of signals. Specifically, with $x(t)$ denoting an energy-dominated homogeneous signal, the expansion in an orthonormal wavelet basis is

$$x(t) = \sum_n \sum_m \bar{x}_m^n \psi_{m,n}(t)$$  

(5.7a)

$$x_m^n = \int_{-\infty}^{\infty} x(t) \psi_{m,n}(t).$$  

(5.7b)

Since $x(t)$ satisfies (5.2) and since $\psi_{m,n}(t)$ satisfies (2.6), it easily follows from (5.7b) that for homogeneous signals

$$x_m^n = \beta^{-m/2} x_0^n$$  

(5.8)

where

$$\beta = 2^{2H+1} = 2^H.$$  

(5.9)

Denoting $x_0^n$ by $q[n]$, (5.7a) then becomes

$$x(t) = \sum_n \sum \beta^{-m/2} q[n] \psi_{m,n}(t).$$  

(5.10)

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

Let us now specifically choose the ideal bandpass wavelet basis, whose basis functions we denote by

$$\hat{\psi}_m^n(t) = 2^{m/2} \hat{\psi}(2^m t - n)$$  

(5.11)

where $\hat{\psi}(t)$ is the ideal bandpass wavelet whose Fourier transform is given by (2.7). If we sample the output $x_0(t)$ of the filter in Definition 5.1 at unit rate, we obtain the sequence $\hat{q}[n] = x_0^n$, where $\hat{q}[n]$ denotes the coefficients of expansion of $x(t)$ in terms of the ideal bandpass wavelet basis. Since $x_0(t)$ has the orthonormal expansion

$$x_0(t) = \sum_n \hat{q}[n] \hat{\psi}_0^n(t)$$  

(5.12)

we have

$$\int_{-\infty}^{\infty} x_0^n(t) dt = \sum_n \hat{q}^2[n].$$  

(5.13)

Consequently, a homogeneous function is energy-dominated if and only if its generating sequence in terms of the ideal bandpass wavelet basis has finite energy, i.e.,

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

A convenient inner product between two energy-dominated homogeneous signals $f(t)$ and $g(t)$ can be defined as

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f_0(t) g_0(t) dt$$  

where the signals $f_0(t)$ and $g_0(t)$ are the responses of the bandpass filter (5.3) to $f(t)$ and $g(t)$, respectively. Exploiting (5.12) we may more conveniently express this inner product in terms of the respective generating sequences of $f(t)$ and $g(t)$ under the bandpass wavelet basis, as

$$\langle f, g \rangle = \sum_n \hat{a}[n] \hat{b}[n].$$  

(5.14)

With this inner product, $E^H$ constitutes a Hilbert space and the induced norm on $E^H$ is

$$\|x\|_\psi^2 = \int_{-\infty}^{\infty} x_0^n(t) dt = \sum_n \hat{q}^2[n].$$  

(5.15)

One can readily construct "self-similar" bases for $E^H$. Indeed, the ideal bandpass wavelet basis (5.11) immediately provides an orthonormal basis for $E^H$. In particular, for any $x(t) \in E^H$, we have the synthesis/analysis pair

$$x(t) = \sum_n \hat{q}[n] \hat{\psi}_H^n(t)$$  

(5.16a)

$$\hat{q}[n] = (x, \hat{\psi}_H^n)_\psi$$  

(5.16b)
where
\[ \hat{\psi}^H_n(t) = \sum_m \beta^{-m/2} \psi_n^m(t). \] (5.17)

One can readily verify that the basis functions (5.17) are self-similar, orthogonal, and have unit norm.

The fact that the ideal bandpass basis is unrealizable means that (5.16) is not a practical mechanism for synthesizing or analyzing homogeneous signals. However, more practical wavelet bases are equally suitable for defining an inner product for the Hilbert space \( EH \). In fact, we now 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.

We begin by noting that not every orthonormal wavelet basis can be used to define inner products for \( EH \). In order to determine which orthonormal wavelet bases can be used for this purpose, we must determine for which wavelets \( \psi(t) \)

That is, we seek conditions on a wavelet basis such that the sequence

\[ q[n] = \int_{-\infty}^{\infty} x(t) \psi_n^0(t) \, dt \in l^2(\mathbb{Z}) \Leftrightarrow x(t) = \sum_m \sum_n \beta^{-m/2} q[n] \psi_n^m(t) \in EH. \]

That is, we seek conditions on a wavelet basis such that the sequence

\[ q[n] = \int_{-\infty}^{\infty} x(t) \psi_n^0(t) \, dt \]

has finite energy whenever the homogeneous signal \( x(t) \) is energy-dominated, and simultaneously such that the homogeneous signal

\[ x(t) = \sum_m \sum_n \beta^{-m/2} q[n] \psi_n^m(t) \]

is energy-dominated whenever the sequence \( q[n] \) has finite energy. Our main result is presented in terms 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.,

\[ \psi^{(r)}(0) = 0, \quad r = 0, 1, \ldots, R - 1 \] (5.18)

and let

\[ x(t) = \sum_m \sum_n \beta^{-m/2} q[n] \psi_n^m(t) \]

be a bihomogeneous signal whose degree \( \gamma \) is such that \( \gamma = \log_2 \beta = 2H + 1 \) satisfies \( 0 < \gamma < 2R - 1 \). Then \( x(t) \) is energy-dominated if and only if \( q[n] \) has finite energy.

This theorem implies that for our Hilbert space \( EH \) we may choose from among a large number of inner products whose induced norms are all equivalent. In particular, for any wavelet \( \psi(t) \) with sufficiently many vanishing moments, we may define the inner product between two functions \( f(t) \) and \( g(t) \) in \( EH \) whose generating sequences are \( a[n] \) and \( b[n] \), respectively, as

\[ (f, g)_\psi = \sum_n a[n] b[n]. \] (5.19)

Of course, this collection of inner products is almost surely not exhaustive. Even for wavelet-based inner products, Theorem 5.3 asserts only that the vanishing moment condition is sufficient to ensure that the inner product generates an equivalent norm. It seems unlikely that the vanishing moment condition is a necessary condition.

The wavelet-based norms for \( EH \) constitute a highly convenient and practical collection from which to choose in applications involving the use of homogeneous signals. Indeed, each associated wavelet-based inner product leads immediately to an orthonormal self-similar basis for \( EH \): if \( x(t) \in EH \), then

\[ x(t) = \sum_n q[n] \theta_n^H(t) \quad (5.20a) \]

\[ q[n] = \langle x, \theta_n^H \rangle_\psi \quad (5.20b) \]

where, again, the basis functions

\[ \theta_n^H(t) = \sum_m \beta^{-m/2} \psi_n^m(t) \] (5.21)

are all self-similar, mutually orthogonal, and have unit norm.

As an example for the case \( H = 0 \), Fig. 5.1 depicts the self-similar basis functions \( \theta_0^0(t) \), \( \theta_1^0(t) \), \( \theta_2^0(t) \), and \( \theta_3^0(t) \) corresponding to the Daubechies 5th-order compactly supported wavelet basis. These functions were generated by evaluating the summation (5.21) 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 \( x(t) \) yields a different \( q[n] \), though all have finite energy.

It is useful to note that for an arbitrary nonhomogeneous signal \( x(t) \), the sequence

\[ q[n] = \langle x, \theta_n^H \rangle_\psi \]

defines the projections of \( x(t) \) onto \( EH \), so that

\[ x(t) = \int_{-\infty}^{\infty} q[n] \theta_n^H(t) \, dt \]

represents the closest homogeneous signal to \( x(t) \) with respect to the induced norm \( \| \cdot \|_\psi \), i.e.,

\[ x(t) = \arg\min_{y(t) \in EH} \| y - x \|_\psi. \]
In Chapter 6, it will be apparent how such projections arise rather naturally in treating problems of estimation with homogeneous signals.

Finally, we remark that wavelet-based characterizations also 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 (5.10) we get, via some routine algebra,

\[
X(\omega) = \sum_{m} 2^{-mH} \lim_{T \to \infty} \int_{-T}^{T} x(t) e^{-j\omega t} dt \Psi(2^{-m}\omega)Q(2^{-m}\omega)
\]

(5.22)

where \( Q(\omega) \) is the discrete-time Fourier transform of \( q[n] \). This spectrum is to be interpreted in the sense of Theorem 5.2, i.e., \( X(\omega) \) defines the spectral content of the output of a bandpass filter at every frequency \( \omega \) within the passband.

In summary, we have shown that a broad class of wavelet-based norms are equivalent for \( E^H \) in a mathematical sense, and that each of these norms is associated with a particular inner product. An interesting open question concerns whether every equivalent norm for \( E^H \) can be associated with a wavelet basis, in which case the basis functions associated with every orthonormal basis for \( E^H \) could be expressed in terms of some wavelet according to (5.21).

In any case, regardless of whether the collection of inner products we construct is exhaustive or not, they at least constitute a highly convenient and practical collection from which to choose in any given application involving the use of homogeneous signals.

5.3 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 find application as information-bearing waveforms in Chapter 6. The definition and properties closely parallel those for energy-dominated homogeneous signals.

Definition 5.4 A bihomogeneous 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 \( y(t) \) has finite power, i.e.,

\[
\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} y(t) dt < \infty.
\]

The notation \( P^H \) is used to designate the class of power-dominated homogeneous signals of degree \( H \). Moreover, while our definition necessarily includes the energy-dominated signals, which have zero power, insofar as our discussion is concerned they constitute a degenerate case.

Analogous to Theorem 5.2 for the energy-dominated case, we can establish the following theorem describing the spectral properties of power-dominated homogeneous signals.

Theorem 5.5 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

\[
S_y(\omega) = \frac{1}{2T} \left| \int_{-T}^{T} y(t) e^{-j\omega t} dt \right|^2 = \begin{cases} S_x(\omega) & |\omega| < |\omega_x| \\ 0 & \text{otherwise} \end{cases}
\]

(5.23)
Theorem 5.2. Note that since

\[ |w|^{2H+1}S_x(\omega) = |2^k\omega|^{2H+1}S_x(2^k\omega). \] (5.24)

The details of the proof of this theorem are contained in Appendix D.3, although the approach is directly analogous to the proof of its counterpart, Theorem 5.2. Note that since \( S_x(\omega) \) in the theorem does not depend on \( \omega_L \) or \( \omega_U \), this function may be interpreted as the generalized power spectrum of \( x(t) \). Furthermore, the relation (5.24) implies that signals in \( PH \) have a generalized time-averaged power spectrum that is \( 1/f \)-like, i.e.,

\[ S_x(\omega) \sim \frac{1}{|\omega|^\gamma} \] (5.25)

where, via (5.9), \( \gamma = 2H + 1 \).

Theorem 5.5 directly implies that a homogeneous signal \( x(t) \) is power-dominated if and only if its generating sequence \( q[n] \) in the ideal bandpass wavelet basis has finite power, i.e.,

\[ \lim_{L \to \infty} \frac{1}{2L+1} \sum_{n=-L}^{L} q[n]^2 < \infty. \]

Similarly we can readily deduce from the results of Section 5.2 that, in fact, for any orthonormal wavelet basis with \( R > H + 1 \) vanishing moments, 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 when we use (5.20a) with such wavelets to synthesize a homogeneous signal \( x(t) \) using an arbitrary finite power sequence \( q[n] \), we are assured that \( x(t) \in PH \). Likewise, when we use (5.20b) to analyze any signal \( x(t) \in PH \), we are assured that \( q[n] \) has finite power.

Some general remarks are appropriate at this point in the discussion. 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 the statistically self-similar \( 1/f \) processes of corresponding degree \( H \), whose power spectra are also of the form (5.25) with \( \gamma = 2H + 1 \) [cf. (3.6)]. In turn, this suggests that, when defined, power-dominated homogeneous signals and \( 1/f \) processes of the same degree also have identical Hausdorff-Besicovitch dimensions [4]. 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 alongside a power-dominated homogeneous signal of the same degree whose generating sequence has been taken from a white random process. We stress, however, that in Fig. 5.2(a), the self-similarity of the

Sec. 5.3 Power-Dominated Homogeneous Signals

\( 1/f \) process is statistical; i.e., a typical sample function does not satisfy (5.2) but its autocorrelation function does. In Fig. 5.2(b), the self-similarity of the homogeneous signal is deterministic. In fact, while the wavelet coefficients of homogeneous signals are identical from scale to scale to within an amplitude factor, i.e.,

\[ x_m^n = \beta^{-m/2}q[n], \]

recall from Chapter 3 that the wavelet coefficients of \( 1/f \) processes have only the same second-order statistics from scale to scale to within an amplitude factor, i.e.,

\[ E[x_m^n x_{m+l}^n] = \beta^{-m}p[n-l] \]

for some function \( p[n] \) that is independent of \( m \).

We can quantify the apparent 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 \( \gamma \). 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 5.6 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

\[ x(t) = \sum_{n} \sum_{\infty} j^{-m/2}q[n]\psi_{m}(t) \] (5.26)

using a correlation-ergodic (e.g., Gaussian), zero-mean, stationary white random sequence \( q[n] \) of variance \( \sigma^2 \) has a generalized time-averaged power spectrum of the form

\[ S_x(\omega) = \sigma^2 \sum_{m=0}^{\infty} 2^{-2m}|\Psi(2^{-m}\omega)|^2. \] (5.27)

Note that the time-averaged spectrum (5.27) is identical to the time-averaged spectrum (3.36) for the wavelet-based synthesis of \( 1/f \) processes described in Section 3.3.2. However, we must be careful not to misinterpret this result. It does not suggest that (5.26) 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 3.3.2: 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.
power-dominated homogeneous signals are bounded on any finite interval of the frequency axis that does not include \( \omega = 0 \). However, it is important to appreciate that not all power-dominated homogeneous signals have spectra that are bounded on \( \pi \leq \omega \leq 2\pi \). An interesting subclass of power-dominated homogeneous signals with such unbounded spectra arises, in fact, in our application in Chapter 6. For these signals, \( z(t) \) as defined in Definition 5.4 is periodic, so we refer to this class of power-dominated homogeneous signals 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. Furthermore, their power spectra consist of impulses 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 (5.10) in which the generating sequence \( q[n] \) is periodic.

### 5.4 Discrete-Time Algorithms for Homogeneous Signals

Orthonormal wavelet representations provide some useful insights into homogeneous signals. For instance, because the sequence \( q[n] \) is replicated at each scale in the representation (5.10) of a homogeneous signal \( x(t) \), the detail signals

\[
D_n x(t) = \beta^{-n/2} \sum_n q[n] \psi_n^m(t)
\]

representing \( q[n] \) modulated into a particular octave band are simply time-dilated versions of one another, to within an amplitude factor. The corresponding time-frequency portrait of a homogeneous signal is depicted in Fig. 5.3, from which the scaling properties are apparent. For purposes of illustration, the signal in this figure has degree \( H = -1/2 \) (i.e., \( \beta = 1 \)), which corresponds to the case in which \( q[n] \) is scaled by the same amplitude factor in each octave band. As always, 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 do for \( 1/f \) processes as discussed in Chapters 3 and 4. 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 \( q[n] \) which corresponds to the coefficients of the expansion of \( x(t) \) in an orthonormal basis \( \{ \psi_n(t) \} \) for \( E^H \). When the \( \theta_n^H(t) \) are derived from a wavelet
basis according to (5.21), another useful discrete-time representation for \( x(t) \) is available, which we now discuss.

Consider the coefficients \( a^m_n \) characterizing the resolution-limited approximation \( \phi_m(x(t)) \) of a homogeneous signal \( x(t) \) with respect to a particular wavelet-based multiresolution signal analysis. Since these coefficients are the projections of \( x(t) \) onto dilations and translations of the scaling function \( \phi(t) \) according to (2.13), it is straightforward to verify that they, too, are identical at all scales to within an amplitude factor, i.e.,

\[
a^m_n = \beta^{-m/2} a^0_n. \tag{5.28}
\]

Consequently, the sequence \( a^m_n \) is an alternative discrete-time characterization of \( x(t) \), since knowledge of it is sufficient to reconstruct \( x(t) \) to arbitrary accuracy. For convenience, we refer to \( a^m_n \) as the characteristic sequence and denote it as \( p[n] \). As is true for the generating sequence, the characteristic sequence associated with \( x(t) \) depends upon the particular multiresolution analysis used; distinct multiresolution signal analyses generally yield different characteristic sequences for any given homogeneous signal. In what follows, we restrict our attention to multiresolution analyses whose basic wavelet meets the vanishing moment conditions of Theorem 5.5.

The characteristic sequence \( p[n] \) is associated with a resolution-limited approximation to the corresponding homogeneous signal \( x(t) \). Specifically, \( p[n] \) represents unit-rate samples of the output of the filter, driven by \( x(t) \), whose frequency response is \( \Phi'(|\omega|) \), the complex conjugate of the Fourier transform of the scaling function. Because frequencies in a neighborhood of the spectral origin, where the spectrum of \( x(t) \) diverges, are passed by such a filter, \( p[n] \) often has infinite energy or, worse, infinite power, even when the generating sequence \( q[n] \) has finite energy.

The characteristic sequence can, in fact, be viewed as a discrete-time homogeneous signal, and a theory can be developed following an approach directly analogous to that used in Sections 5.2 and 5.3 for the case of continuous-time homogeneous signals. The characteristic sequence satisfies the discrete-time self-similarity relation

\[
j^{1/2} p[n] = \sum_k h[k - 2n] p[k], \tag{5.29}\]

which is readily obtained by substituting for \( a^m_n \) in the DWT analysis equation (2.21a) using (5.28). Indeed, as depicted in Fig. 5.4, (5.29) 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 recover an amplitude-scaled version of \( p[n] \). Although characteristic sequences are, in an appropriate sense, "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 the sequence \( p[n] \) as energy-dominated in the former case, and power-dominated in the latter case. In fact, when the output of such a highpass filter is downsampled at rate two, we recover the characteristic sequence \( q[n] \) associated with the expansion of \( x(t) \) in the corresponding wavelet basis, i.e.,

\[
j^{1/2} q[n] = \sum_k g[k - 2n] p[k]. \tag{5.30}\]

This can be readily verified by substituting for \( a^m_n \) and \( x^m_n \) in the DWT analysis equation (2.21b) using (5.28) and (5.8), and by recognizing that \( a^m_n = p[n] \) and \( x^m_n = q[n] \).

From a different perspective, (5.30) provides a convenient mechanism for obtaining the representation for a homogeneous signal \( x(t) \) in terms of its

\[\text{Relations of this type may be considered discrete-time counterparts of the dilation equations considered by Strang [26].}\]
generating sequence $q[n]$ from one in terms of its corresponding characteristic sequence $p[n]$, i.e.,

$$p[n] \rightarrow q[n].$$

To obtain the reverse mapping

$$q[n] \rightarrow p[n]$$

is less straightforward. For an arbitrary sequence $q[n]$, the associated characteristic sequence $p[n]$ is the solution to the linear equation

$$3^{-1/2}p[n] = \sum_{k} h[n - 2k]p[k] = \sum_{k} g[n - 2k]q[k].$$

(5.31)

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. However, the DWT synthesis algorithm 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

$$p[0][n] = 0$$

$$p[i+1][n] = 3^{1/2} \sum_{k} \{h[n - 2k]p[i][k] + g[n - 2k]q[k]\}.$$  

(5.32a)

(5.32b)

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 $0 < \omega < \pi$. Note that the precomputable quantity

$$q_{s}[n] = \sum_{k} g[n - 2k]q[k]$$

represents the sequence $q[n]$ modulated into essentially the upper half band of frequencies.

Any real application of homogeneous signals can ultimately exploit scaling properties over only a finite range of scales, so that it suffices in practice to modulate $q[n]$ into a finite range of contiguous octave bands. Consequently, only a finite number of iterations of the algorithm (5.32) are required. 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 these signals, as will be apparent in our developments of Chapter 6.

As we conclude this chapter, it is worth mentioning that there may be useful connections to be explored between the self-similar signal theory described here and the work of Barnsley [92] on deterministically self-affine one-dimensional and multi-dimensional signals. Malassenet and Mersereau [93], for example, suggest that these so-called "iterated function systems" have efficient representations in terms of wavelet bases as well.

---

**5.5 SUMMARY**

In this chapter we focused on fractal signals characterized by a deterministic scaling relation. We showed that these "homogeneous" signals, in contrast to the fractal random processes described in Chapter 3, have the property that the waveforms themselves remain invariant to within an amplitude factor under arbitrary scaling of the time axis.

We then introduced and developed a new and richer generalized family of homogeneous signals defined in terms of a dyadic scale-invariance property. When necessary to avoid confusion with traditional homogeneous signals, we specifically referred to this broader family as bihomogeneous signals. Motivated by our interest in using these signals as modulating waveforms in some communication applications in the next chapter, we proceeded to develop some of their important properties and representations.

We began by distinguishing between two classes: energy-dominated and power-dominated, and then developed their spectral properties. We then showed that the use of wavelet basis expansions leads to constructions of powerful orthonormal self-similar bases for homogeneous signals. From this perspective, we saw that wavelet representations play as natural and important a role in the representation of these signals as they did for the 1/f processes developed in Chapter 3.

In the latter portion of the chapter, we exploited the discrete wavelet transform algorithm to derive highly efficient discrete-time algorithms for both synthesizing and analyzing homogeneous signals using these representations. As we will see, these algorithms play an important role in the transmitters and receivers of the communication system we explore in the next chapter.
Fractal Modulation

6.1 INTRODUCTION

There are a number of interesting potential applications for the homogeneous signal theory developed in the Chapter 5. In this chapter, we focus on a particular example as an indication of the direction that some applications may take. In particular, we explore the use of homogeneous signals as modulating waveforms in a communication system [91]. Beginning with an idealized but general channel model, we demonstrate that the use of homogeneous waveforms in such channels is both natural and efficient, and leads to a novel multirate diversity strategy in which data is transmitted simultaneously at multiple rates.

Our problem involves the design of a communication system for transmitting a continuous- or discrete-valued data sequence over a noisy and unreliable continuous-amplitude, continuous-time channel. We must therefore design a modulator at the transmitter that embeds the data sequence \( q[n] \) into a signal \( x(t) \) to be sent over the channel. At the receiver, a demodulator must be designed for processing the distorted signal \( r(t) \) from the channel to extract an optimal estimate of the data sequence \( q[n] \). The overall system is depicted in Fig. 6.1.

The particular channel we consider has the characteristic that it is "open" for some time interval \( T \), during which it has a particular bandwidth \( W \) and signal-to-noise ratio (SNR). This rather basic channel model is a useful one for a variety of settings, and in particular it 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 methodologies for designing an efficient and reliable communication system. However, we restrict our attention to the case in which both the bandwidth and duration parameters are either unknown or not available to the transmitter. This case, by contrast, has received comparatively less attention in the communications literature, although it arises rather naturally in a range of both point-to-point and multiuser communication scenarios involving, for example, jammed and fading channels, multiple access channels, covert and low probability of intercept (LPI) communication, and broadcast communication to disparate receivers.

In designing a suitable communication system for such channels, we require that the following key performance characteristics be satisfied:

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, i.e., \( z(t) = 0 \).
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.

The first of these requirements implies that, at least in principle, we ought to be able to recover \( q[n] \) using arbitrarily narrow receiver bandwidth given sufficient duration, or, alternatively, from an arbitrarily short duration segment given sufficient bandwidth. The second requirement implies that we ought to be able to obtain better estimates of \( q[n] \) the longer a receiver is able to listen, or the greater the bandwidth it has available. Consequently, the modulation must contain redundancy or diversity of a type that can be exploited for the purposes of improving the reliability of the transmission. As we demonstrate, the use of homogeneous signals for transmission appears to be rather naturally suited to fulfilling both these system requirements.

The minimum achievable duration-bandwidth threshold in such a system is a measure of the efficiency of the modulation. Actually, because the duration-bandwidth threshold \( T \times W \) is a function of the length \( L \) of the data sequence, it is more convenient to transform the duration constraint \( T \) into a symbol rate constraint \( R = L/T \) and phrase the discussion in terms of a rate-bandwidth threshold \( R/W \) that is independent of sequence length. Then the maximum achievable rate-bandwidth threshold constitutes the spectral efficiency of the modulation, which we denote by \( \eta \). The spectral efficiency of a transmission scheme using bandwidth \( W \) is, in fact, defined as

\[
\eta = R_{\text{max}} / W
\]

where \( R_{\text{max}} \) is the maximum rate at which perfect communication is possible in the absence of noise. Hence, the higher the spectral efficiency of a scheme, the higher the rate that can be achieved for a given bandwidth, or, equivalently, the smaller the bandwidth that is required to support a given rate.

When the available channel bandwidth is known \textit{a priori}, a reasonably spectrally efficient, if impractical, modulation of a data sequence \( q[n] \) involves expanding the sequence in terms of an ideally bandlimited orthonormal basis. Specifically, with \( W_0 \) denoting the channel bandwidth, a transmitter produces

\[
x(t) = \sum_n q[n] \sqrt{W_0} \sin(W_0t - n)
\]

where

\[
sinc(t) = \begin{cases} 
1 & t = 0 \\
\sin \pi t / \pi t & \text{otherwise}
\end{cases}
\]

In the absence of noise, a (coherent) receiver can, in principle, recover \( q[n] \) from the projections

\[
q[n] = \int_{-\infty}^{\infty} x(t) \sqrt{W_0} \sin(W_0t - n) \, dt
\]

which can be implemented as a sequence of filter-and-sample operations. Since this scheme achieves a rate of \( R = W_0 \) symbols/sec using the double-sided bandwidth of \( W = W_0 \) Hz, it is characterized by a spectral efficiency of

\[
\eta_0 = 1 \text{ symbol/sec/Hz}.
\]

Because the transmitter is assumed to have perfect knowledge of the rate-bandwidth characteristics of the channel, this approach is viable only for those point-to-point channels in which there exists a feedback path from receiver to transmitter. However, we consider a more general case in which we have either a broadcast channel (i.e., a scenario with a single transmitter and multiple receivers), or a point-to-point channel in which no such feedback path is available. In these cases, the approach outlined above does not constitute a viable solution to our communications problem. Indeed, in order to accommodate a decrease in available channel bandwidth, the transmitter would have to be accordingly reconfigured by decreasing the parameter \( W_0 \). Similarly, for the system to maintain a spectral efficiency of \( \eta_0 = 1 \) when the available channel bandwidth increases, the transmitter must be reconfigured by correspondingly increasing the parameter \( W_0 \). Nevertheless, while not a solution to the problem of communication without a feedback path, the perfect-feedback solution provides a useful performance baseline in evaluating the solution we ultimately develop for this problem. In the sequel, we therefore refer to this as our benchmark modulation scheme.

A viable solution to the problem of interest requires a modulation strategy that maintains its spectral efficiency over a broad range of rate-bandwidth combinations using a fixed transmitter configuration. A rather natural strategy of this type arises out of the concept of embedding the data to be transmitted into a homogeneous signal. Due to the fractal properties of the transmitted signals, we refer to the resulting scheme as "fractal modulation."

### 6.2 Transmitter Design: Modulation

To embed a finite-power sequence \( q[n] \) into a bihomogeneous waveform \( x(t) \) of degree \( H \), it suffices to consider using \( q[n] \) as the coefficients of an expansion in terms of a wavelet-based orthonormal self-similar basis of degree \( H \), i.e.,

\[
x(t) = \sum_n q[n] \delta^H_n(t)
\]

where the basis functions \( \delta^H_n(t) \) are constructed according to (5.21). When the basis is derived from the ideal bandpass wavelet, as we will generally
assume in our analysis, the resulting waveform $x(t)$ is a power-dominated homogeneous signal whose idealized time-frequency portrait has the form depicted in Fig. 5.3. Consequently, we may view this as a *multirate modulation* of $q[n]$ where in the $m$th frequency band $q[n]$ is modulated at rate $2^n$ using a double-sided bandwidth of $2^m$ Hz. Furthermore, the energy per symbol used in successively higher bands scales by $\beta = 2^{2m+1}$. Using a suitably designed receiver, $q[n]$ can, in principle, be recovered from $x(t)$ at an arbitrary rate $2^n$ using a baseband bandwidth of $2^{n+1}$ Hz. Consequently, this modulation has a spectral efficiency of

$$\eta_F = \frac{1}{2} \text{ symbol/sec/Hz}.$$  

We emphasize that in accordance with our channel model of Fig. 6.2, it is the baseband bandwidth that is important in defining the spectral efficiency since it defines the highest frequency available at the receiver.

While the spectral efficiency of this modulation is half that of the benchmark scheme (6.1), this loss in efficiency is, in effect, the price paid to enable a receiver to use any of a range of rate-bandwidth combinations in demodulating the data. Fig. 6.3 illustrates the rate-bandwidth tradeoffs available to the receiver. In the absence of noise the receiver can, in principle, perfectly recover $q[n]$ using rate-bandwidth combinations lying on or below the solid curve. The stepped character of this curve reflects the fact that only rates of the form $2^n$ can be accommodated, and that full octave increases in bandwidth are required to enable $q[n]$ to be demodulated at successively higher rates. For reference, the performance of the benchmark modulation is superimposed on this plot using a dashed line. We emphasize that in contrast to fractal modulation, the transmitter in the benchmark scheme requires perfect knowledge of the rate-bandwidth characteristics of the channel.

Although it considerably simplifies our analysis, the use of the ideal bandpass wavelet to synthesize the orthonormal self-similar basis in our modulation strategy is impractical due to the poor temporal localization in this wavelet. However, we may, in practice, replace the ideal bandpass wavelet with one having not only comparable frequency domain characteristics and better temporal localization, but sufficiently many vanishing moments to ensure that the transmitted waveform is power-dominated as well. Fortunately, there are many suitable wavelets from which to choose, among which are those due to Daubechies [12]. When such wavelets are used, the exact spectral efficiency of the modulation depends on the particular definition of bandwidth employed. Nevertheless, using any reasonable definition of bandwidth, we would expect to be able to achieve, in practice, a spectral efficiency close to $(1/2)$ symbols/sec/Hz with this modulation, and, as a result, we assume $\eta_F \approx 1/2$ in subsequent analysis.

Another apparent problem with fractal modulation as initially proposed is that it requires infinite transmitter power. Indeed, as Fig. 5.3 illustrates,
the wavelet basis. The second stage then consists of a discrete- to continuous-
time transformation in which $p^{[M]}[n]$ is modulated into the continuous-time
frequency spectrum via the appropriate scaling function according to
\[ x(t) = \sum_n p^{[M]}[n] \phi_n^M(t) = \sum_n p^{[M]}[n] 2^M \phi(2^M t - n). \]

It is important to point out that because a batch-iterative algorithm is em-
ployed, 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 possible that memory-efficient implementations may be
constructed as well.

The transmission of finite length sequences using fractal modulation
requires some basic modifications to the scheme. In fact, as initially pro-
posed, fractal modulation is rather inefficient in this case, in essence because
successively higher frequency bands are increasingly underutilized. In par-
ticular, we note from the time-frequency portrait in Fig. 5.3 that if $q[n]$ has
finite length, e.g.,
\[ q[n] = 0, \quad n < 0, n > L - 1, \]
then the $m$th band completes its transmission of $q[n]$ and goes idle in half the
time it takes the $(m-1)$st band, and so forth. However, finite length messages
may be accommodated rather naturally and efficiently by modulating their
periodic extensions $q[n \mod L]$ thereby generating a transmitted waveform
\[ x(t) = \sum_n q[n \mod L] \theta_n^M(t) \]
which constitutes a periodicity-dominated homogeneous signal of the type
discussed in Section 5.3. If we let
\[ q = \{ q[0], q[1], \ldots, q[L-1] \} \]
denote the data vector, then the time-frequency portrait associated with this
signal is shown in Fig. 6.4. Using this enhancement of fractal modulation,
we not only maintain our ability to make various rate-bandwidth tradeoffs
at the receiver, 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 more generally 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 data transmission on
a block-by-block basis.

The final aspect of fractal modulation that remains to be considered
in this section concerns the specification of the parameter $H$. While $H$ has
no effect on the spectral efficiency of fractal modulation, it does affect the
power efficiency of the scheme. Indeed, it controls the relative power dis-
tribution between frequency bands and, hence, the overall transmitted power

![Figure 6.4. 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.](image-url)
As we discussed in Chapter 3, the class of 1/f processes includes not only classical white Gaussian noise \( (H = -1/2) \) and Brownian motion \( (H = 1/2) \), but, more generally, a range of rather prevalent nonstationary noises that exhibit strong long-term statistical dependence.

In this section, we have developed a modulation strategy that satisfies the first of the two system requirements described at the outset of Chapter 6. We now turn our attention to the problem of designing optimal receivers for fractal modulation, and, in the process, we will see that fractal modulation also satisfies the second of our key system requirements.

### 6.3 RECEIVER DESIGN: DEMODULATION

Consider the problem of recovering a finite length message \( q[n] \) from band-limited, time-limited, and noisy observations \( r(t) \) of the transmitted waveform \( x(t) \) consistent with our channel model of Fig. 6.2. We assume that the noise \( z(t) \) is a Gaussian 1/f process of degree \( H_z = H \), and that the degree \( H_x \) of the homogeneous signal \( x(t) \) has been chosen according to our spectral-matching rule, i.e.,

\[
H_x = H_z = H. \tag{6.3}
\]

We remark at the outset that if it is necessary that the transmitter measure \( H_x \) in order to perform this spectral matching, the robust and efficient parameter estimation algorithms for 1/f processes developed in Chapter 4 may be conveniently exploited.

Depending on the nature of the message being transmitted, there are a variety of different optimization criteria from which to choose in designing a suitable receiver. As representative examples, we consider two cases. In the first, the transmitted message is a digital data stream and we focus on the optimal decoding of each bit can be described in terms of a binary hypothesis test on the set of available observation coefficients \( r \). Denoting by \( q[n] \) the binary-valued sequence available at the receiver is

\[
K = \sum_{m=M_L}^{M_U} 2^{m-M_U} = 2^{M_U-M_L+1} - 1 \tag{6.6}
\]

doing more practical wavelets are used.

The duration-bandwidth characteristics of the channel in general affect which observation coefficients \( r_n^m \) may be accessed. In particular, if the channel is bandlimited to \( 2^{M_U} \) Hz for some integer \( M_U \), this precludes access to the coefficients at scales corresponding to \( m > M_U \). Simultaneously, the duration-constraint in the channel results in a minimum allowable decoding rate of \( 2^{M_U} \) symbols/sec for some integer \( M_L \), which precludes access to the coefficients at scales corresponding to \( m < M_L \). As a result, the collection of coefficients available at the receiver is

\[
K = \{ r_n^m, m \in M, n \in N(m) \}
\]

where

\[
M = \{ M_L, M_L + 1, \ldots, M_U \} \tag{6.5a}
\]

\[
N(m) = \{ 0, 1, \ldots, L 2^{m-M_U} - 1 \}. \tag{6.5b}
\]

This means that we have available

\[
K = \sum_{m=M_L}^{M_U} 2^{m-M_U} = 2^{M_U-M_L+1} - 1
\]

doing more practical wavelets are used.

The duration-bandwidth characteristics of the channel in general affect which observation coefficients \( r_n^m \) may be accessed. In particular, if the channel is bandlimited to \( 2^{M_U} \) Hz for some integer \( M_U \), this precludes access to the coefficients at scales corresponding to \( m > M_U \). Simultaneously, the duration-constraint in the channel results in a minimum allowable decoding rate of \( 2^{M_U} \) symbols/sec for some integer \( M_L \), which precludes access to the coefficients at scales corresponding to \( m < M_L \). As a result, the collection of coefficients available at the receiver is

\[
K = \{ r_n^m, m \in M, n \in N(m) \}
\]

where

\[
M = \{ M_L, M_L + 1, \ldots, M_U \} \tag{6.5a}
\]

\[
N(m) = \{ 0, 1, \ldots, L 2^{m-M_U} - 1 \}. \tag{6.5b}
\]

This means that we have available

\[
K = \sum_{m=M_L}^{M_U} 2^{m-M_U} = 2^{M_U-M_L+1} - 1
\]

doing more practical wavelets are used.

The duration-bandwidth characteristics of the channel in general affect which observation coefficients \( r_n^m \) may be accessed. In particular, if the channel is bandlimited to \( 2^{M_U} \) Hz for some integer \( M_U \), this precludes access to the coefficients at scales corresponding to \( m > M_U \). Simultaneously, the duration-constraint in the channel results in a minimum allowable decoding rate of \( 2^{M_U} \) symbols/sec for some integer \( M_L \), which precludes access to the coefficients at scales corresponding to \( m < M_L \). As a result, the collection of coefficients available at the receiver is

\[
K = \{ r_n^m, m \in M, n \in N(m) \}
\]

where

\[
M = \{ M_L, M_L + 1, \ldots, M_U \} \tag{6.5a}
\]

\[
N(m) = \{ 0, 1, \ldots, L 2^{m-M_U} - 1 \}. \tag{6.5b}
\]

This means that we have available

\[
K = \sum_{m=M_L}^{M_U} 2^{m-M_U} = 2^{M_U-M_L+1} - 1
\]

The optimal decoding of each bit can be described in terms of a binary hypothesis test on the set of available observation coefficients \( r \). Denoting by \( H_f \) the hypothesis in which \( q[n] = +\sqrt{E_0} \), and by \( H_0 \) the hypothesis in which
$q[n] = -\sqrt{E_0}$, we may construct the likelihood ratio test for the optimal decoding of each symbol $q[n]$. The derivation is particularly straightforward because of the fact that, in accordance with the wavelet-based models for $1/f$ processes developed in Chapter 3, under each hypothesis the $z_n^m$ in (6.4) may be modeled as independent zero-mean Gaussian random variables with variances

$$\text{var} z_n^m = \sigma_z^2 \beta^{-m}$$

(6.8)

for some variance parameter $\sigma_z^2 > 0$. Consequently, given equally likely hypotheses (i.e., a random bit stream) the likelihood ratio test readily reduces to the test

$$H_1 \quad \ell \geq 0, \quad H_0$$

where

$$\ell = \sum_{m=M_t}^{M_t} \beta^{m/2} \sum_{i=0}^{2^m-M_t-1} r_{n+4K}^m$$

(6.9)

is a sufficient statistic.

Before turning to a discussion of the resulting performance, it should be emphasized that, as in the case of the transmitter, the receiver has a convenient, computationally efficient, hierarchical implementation based on the DWT. Specifically, assuming $r(t)$ to be bandlimited to resolution $2^{M_t}$, it may be sampled at rate $2^{M_t}$, then successively filtered and downsampled to level $m = M_t$ according to the wavelet decomposition tree of Fig. 2.6(a). To produce the sufficient statistic $\ell$, at each level $m$ the terms from the detail sequence $t_n^m$ corresponding to the same value of the $q[n]$ are collected together, weighted by the factor $\beta^{m/2}$, and accumulated with the weighted $r_n^m$ from previous stages. Again, however, this is a batch algorithm, and while computationally efficient, this implementation may be less efficient in terms of storage requirements.

**Performance**

Since the statistic $\ell$ is Gaussian under each hypothesis, the performance associated with this optimal receiver is straightforward to derive. In particular, since

$$E[\ell|H_1] = -E[\ell|H_0] = \sqrt{E_0}K$$

$$\text{var}[\ell|H_1] = \text{var}[\ell|H_0] = \sigma_z^2 K$$

and since

$$\Pr(\ell > 0) = \Pr(\ell < 0)$$

the bit-error probability can be expressed as

$$\Pr(\varepsilon) = \Pr(\ell > 0|H_0) = Q \left( \sqrt{K \sigma_z^2} \right)$$

(6.10)

where $Q(\cdot)$ is again defined by (4.32), and where $\sigma_z^2$ is the SNR in the channel, i.e.,

$$\sigma_z^2 = \frac{E_0}{\sigma_z^2}$$

Substituting for $K$ in (6.10) via (6.7) we can rewrite this error probability in terms of the channel rate-bandwidth ratio as

$$\Pr(\varepsilon) = Q \left( \sqrt{\frac{\sigma_z^2}{\eta_f R/W - 1}} \right)$$

(6.11)

where, again, $\eta_f \approx 1/2$. Note that, as we would anticipate with the strategy, the performance of fractal modulation is independent of the spectral exponent of the noise process when we use our spectral matching procedure.

To establish a performance baseline, we also evaluate a modified version of our benchmark modulation in which we incorporate repetition coding, i.e., in which we add redundancy by transmitting each sample of the message sequence $K$ times in succession. This comparison scheme is not particularly power efficient both because signal power is distributed uniformly over the available bandwidth irrespective of the noise spectrum, and because much more effective coding schemes can be used with channels of known bandwidth [96] [97]. Nevertheless, with these caveats in mind, such comparisons do lend some insight into the relative power efficiency of fractal modulation.

In our modified benchmark modulation, incorporating redundancy reduces the effective decoding rate per unit bandwidth by a factor of $K$, i.e.,

$$\frac{R}{W} = \frac{\eta_0}{K}$$

(6.12)

where $\eta_0$ is the efficiency of the modulation without coding, i.e., unity. When the channel adds stationary white Gaussian noise, for which $H = -1/2$, the optimum receiver for this scheme 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 [97]

$$\Pr(\varepsilon) = Q \left( \sqrt{K \sigma_z^2} \right) = Q \left( \sqrt{\frac{\sigma_z^2}{R/W}} \right)$$

(6.13)

where the last equality results from substituting for $K$ via (6.12).

Comparing (6.13) with (6.11), we note that since $\eta_0 \approx 2\eta_f$, the asymptotic bit-error performances of fractal modulation and the benchmark scheme
are effectively equivalent for $R/W \ll \eta_r$, as is illustrated in Fig. 6.5. In Fig. 6.5(a), $P(e)$ is shown as a function of $R/W$ at a fixed SNR of 0 dB ($\sigma_r^2 = 1$), while in Fig. 6.5(b), $P(e)$ is shown as a function of SNR at a fixed $R/W = 0.125$ bits/sec/Hz. Both these plots reveal strong thresholding behavior whereby the error probability falls off dramatically at high SNR and low $R/W$. It is important to emphasize that comparisons between the two schemes are meaningful only for the case in which the noise has parameter $H = -1/2$, corresponding to the case of stationary white Gaussian noise. For other values of $H$, the performance of the benchmark modulation is not only difficult to evaluate, but necessarily poor as well because of inefficient distribution of power among frequencies.

6.3.2 Demodulation of Analog Data

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

$$\text{var} q[n] = \sigma_q^2,$$

and develop a receiver yielding the minimum mean-square error (MSE) estimate of $q[n]$ based on our corrupted observations $r(t)$.

We proceed in a manner analogous to that described in Section 6.3.1 for the case of digital data. In particular, we first project our observations onto the ideal bandpass wavelet basis from which $x(t)$ was synthesized, so that our observations may again be expressed in the form (6.4). Given the set of $K$ accessible observation coefficients $r_m$ specified by (6.5) with (6.6), we readily obtain that the optimum estimates of $q[n]$ are of the form

$$\hat{q}[n] = \frac{\ell}{\ell + 1} \sum_{m=1}^{K} \sum_{n=1}^{N} \frac{r_m}{\sigma_q^2}$$

where $\ell$ is the sufficient statistic (6.9) and where $\sigma_q^2$ is the SNR in the channel, defined by

$$\sigma_q^2 = \frac{\sigma_q^2}{\sigma_r^2}$$

Evidently, the sufficient statistic $\ell$ defined in (6.9) plays a key role in the demodulation of both digital and analog data, and in fact its calculation dominates the computational complexity of the receiver. However, we emphasize that, as was discussed in Section 6.3.1, that $\ell$ can be obtained via a computationally efficient algorithm by exploiting the DWT.

In general, and as we would expect, the optimum estimate represents a blend of a priori information about $q[n]$, and information obtained from the observations. At high SNR ($\sigma_q^2 \gg 1/K$), the a priori information is essentially ignored, and the resulting estimator specializes to the maximum likelihood

![Figure 6.5](image-url)
estimator. At low SNR ($\sigma_n^2 \ll 1/K$), the observations are essentially ignored, and the estimator approaches the *a priori* estimate

$$q[n] = E[q[n]] = 0.$$ 

Finally, we remark that the optimum receiver (6.14) with (6.9) 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 (6.14) can be readily derived as

$$\epsilon^2 = \frac{E[|q[n] - \bar{q}[n]|^2]}{E[|q[n]|^2]} = \frac{\frac{1}{\text{var} q[n]}}{1 + K\sigma_n^2} = \frac{1}{1 + K\sigma_n^2}. \quad (6.15)$$

Generally, it is convenient to substitute for $K$ in (6.15) via (6.7) to get

$$\epsilon^2 = \frac{1}{1 + \sigma_n^2 \left( \frac{2\eta_F}{R/W} - 1 \right)}, \quad (6.16)$$

where $\eta_F \approx 1/2$, and where $R/W \leq \eta_F$ by virtue of our definition of $\eta_F$. From (6.16) we see, then, that for $R/W \ll \eta_F$, the MSE is given asymptotically by

$$\epsilon^2 \sim \frac{1}{\sigma_n^2 \frac{2\eta_F}{R/W}}. \quad (6.17)$$

Note that, as in the case of digital data, the performance (6.16) is independent of the parameter $H$ when we use spectral matching.

For the purposes of comparison, consider the MSE performance of our benchmark modulation with repetition coding in the presence of stationary white Gaussian noise. As in the case of digital data, incorporating redundancy reduces the effective rate-bandwidth ratio by a factor of $K$, yielding (6.12). 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 $\bar{q}[n]$. Hence this $K$-fold redundancy leads to a normalized MSE of

$$\epsilon^2 = \frac{E[|q[n] - \bar{q}[n]|^2]}{E[|q[n]|^2]} = \frac{1}{1 + \sigma_n^2 K} \quad (6.18)$$

where $\sigma_n^2$ 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 (6.12) we get

$$\epsilon^2 = \frac{1}{1 + \sigma_n^2 R/W} \quad (6.19)$$

whenever $R/W \leq \eta_0$. By comparison with (6.17) we see that when $R/W \ll \eta_0$, we have

$$\epsilon^2 \sim \frac{1}{\sigma_n^2 \frac{2\eta_F}{R/W}}. \quad (6.20)$$

which is essentially (6.17) since the achievable $\eta_0$ 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 the benchmark modulation with repetition coding of Fig. 6.6. In Fig. 6.6(a), MSE is shown as a function of $R/W$ at a fixed SNR of 0 dB ($\sigma_n^2 = 1$), while in Fig. 6.6(b), MSE is shown as a function of SNR at a fixed $R/W = 0.125$ 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 are appropriate only for the special case of additive white Gaussian noise channels, we reiterate that the performance of fractal modulation (6.16) 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 the benchmark modulation with repetition coding in the presence 1/f noise.

### 6.4 Summary

In this chapter, we developed a novel and powerful diversity strategy that we referred to as fractal modulation. As developed, fractal modulation constitutes a compelling paradigm for communication over noisy channels of simultaneously uncertain duration and bandwidth. As we discussed, this channel is a rather realistic model for a variety of scenarios encountered in point-to-point, broadcast, and multiple-access communications.

With fractal modulation, we used the orthonormal self-similar basis expansions derived in Chapter 5 to develop an approach for modulating discrete- or continuous-valued information sequences onto homogeneous signals. The result was a modulation scheme in which information was embedded in the transmitted waveform on all time scales.

We then considered the problem of optimum demodulation of such transmissions. Two representative cases were considered. The first involved
minimum probability-of-error demodulation of a digital data stream, while the second involved minimum mean-square error demodulation of an analog data stream.

Of considerable practical interest, we also showed that the use of self-similar bases derived from wavelet bases as developed in Chapter 5 leads directly to computationally efficient discrete-time implementations of the resulting transmitters and receivers for fractal modulation systems.

Our development included an evaluation of several aspects of the performance of this diversity strategy on the channel of interest. In order to provide a performance baseline, our analysis included comparisons to more traditional forms of modulation and diversity.

A number of interesting open questions were not addressed in our development. For example, the extent to which fractal modulation is optimum for the unknown duration-bandwidth channel remains to be explored. Closely related is the issue of how to define a useful notion of capacity for such channels, since the usual Shannon capacity degenerates in this case. It is also possible, though not obvious, that efficient coding techniques [98] can be used in conjunction with fractal modulation to provide a more effective diversity benefit on such channels.

As a final remark, we point out that fractal modulation and its generalizations also have a potentially important role to play in secure communications. In such applications, considerations such as vulnerability to detection, interception, and exploitation by various strategies naturally become important [99]. To meet the particular needs that arise in these scenarios, a variety of basic extensions to fractal modulation can be developed. For example, traditional direct-sequence spread-spectrum type techniques [97] can be combined with fractal modulation in a relatively straightforward manner to further enhance its suitability to such applications.

Figure 6.6. Tradeoffs between error, rate, and bandwidth for fractal modulation with the optimum receiver for noisy analog data. The solid lines represent the performance of fractal modulation, while the dashed lines correspond to the performance of the benchmark modulation with repetition coding. (a) MSE \(e^2\) as a function of Rate/Bandwidth ratio \(R/W\) at 0 dB SNR. (b) MSE \(e^2\) as a function of SNR at \(R/W = 0.125\) symbols/sec/Hz.
Linear Self-Similar Systems

7.1 INTRODUCTION

In preceding chapters we have explored several useful classes of statistically and deterministically self-similar signals that arise in engineering applications. This chapter represents a preliminary investigation into the relationships between self-similar signals and an underlying self-similar system theory. 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 book, and in the process suggests some promising future directions for work in this area.

The self-similar systems we ultimately 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 transforms can be used in their analysis. 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.

Overall, our treatment is rather informal in style, reflecting a conscious effort to emphasize the conceptual themes over mathematical rigor and generality. To facilitate this, our development focuses, for example, on input-output descriptions of systems. Our development begins with a brief summary of some results in the theory of linear time-invariant systems. For more extensive treatments, see, e.g., Oppenheim and Willsky [1], Siebert [2], or Kailath [100].

Linear systems are typically defined as follows. Suppose \( y_1(t), y_2(t), \) and \( y_3(t) \) are the responses of a system \( S \{ \cdot \} \) to arbitrary inputs \( x_1(t), x_2(t), \) and \( x_3(t) \), respectively. Then the system is linear when it satisfies, for any \( a \) and \( b \), the superposition principle

\[
S \{ ax_1(t) + bx_2(t) \} = ay_1(t) + by_2(t).
\]  

(7.1)

Linear systems are often conveniently described in terms of the integral

\[
y(t) = S \{ x(t) \} = \int_{-\infty}^{\infty} x(\tau) \kappa(t, \tau) d\tau,
\]

where \( \kappa(t, \tau) \) is the kernel of the linear system and represents the response of the system at time \( t \) to a unit impulse at time \( \tau \), i.e.,

\[
\kappa(t, \tau) \triangleq S \{ \delta(t - \tau) \}.
\]

7.2 LINEAR TIME-INVARIANT SYSTEMS

An important class of linear systems are those that are also time-invariant. A system is time-invariant when it satisfies, for any constant \( \tau \),

\[
S \{ x(t - \tau) \} = y(t - \tau).
\]  

(7.2)

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 \( \kappa(t, \tau) \) satisfies, for any \( b \),

\[
\kappa(t, \tau) = \kappa(t - b, \tau - b).
\]  

(7.3)

For this class of systems, the kernel has the form

\[
\kappa(t, \tau) = \nu(t - \tau)
\]

where \( \nu(t) \) is the familiar impulse response of the system. Furthermore, the corresponding input-output relation is, of course, described in terms of the usual convolution integral,

\[
y(t) = \int_{-\infty}^{\infty} x(\tau) \nu(t - \tau) d\tau \triangleq x(t) * \nu(t).
\]

The eigenfunctions of LTI systems are complex exponentials of the form \( e^{st} \), from which we get that the Laplace transform

\[
X(s) = \int_{-\infty}^{\infty} x(t) e^{-st} dt
\]
possesses the so-called convolution property; i.e., for signals $x(t)$ and $y(t)$ with Laplace transforms $X(s)$ and $Y(s)$, respectively, we have

$$x(t) * y(t) \leftrightarrow X(s)Y(s).$$

### 7.3 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 literature and pattern recognition literature, and in the broader mathematics literature in connection with the Mellin transform.

To explore these systems, suppose that $y(t)$ is the response of a system $S \{ \cdot \}$ to an arbitrary input $x(t)$. Then a system $S \{ \cdot \}$ is said to be scale-invariant whenever, for any constant $\alpha > 0$,

$$S \{ x(t/\alpha) \} = y(t/\alpha). \quad (7.4)$$

A system satisfying both (7.1) and (7.4) is 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 satisfies

$$\kappa(t, \tau) = a\kappa(at, a\tau) \quad (7.5)$$

for any $a > 0$.

An LSI system is generally characterized in terms of the lagged-impulse response pair

$$\xi_+(t) = S \{ \delta(t - 1) \}, \quad \xi_-(t) = S \{ \delta(t + 1) \}. \quad (7.6a)$$

Indeed, when an input $x(t)$ can be decomposed, except at $t = 0$, as

$$x(t) = \int_{-\infty}^{\infty} x(\tau) \delta \left( \frac{t - \tau}{\tau} - 1 \right) \frac{d\tau}{\tau}$$

$$= \int_{0}^{\infty} x(\tau) \delta \left( \frac{t - \tau}{\tau} - 1 \right) \frac{d\tau}{\tau} - \int_{0}^{\infty} x(-\tau) \delta \left( \frac{t - \tau}{\tau} + 1 \right) \frac{d\tau}{\tau}, \quad (7.7)$$

we can exploit the superposition principle (7.1) together with (7.4) to obtain the following input-output relation

$$y(t) = \int_{0}^{\infty} x(\tau) \xi_+ \left( \frac{t - \tau}{\tau} \right) \frac{d\tau}{\tau} - \int_{0}^{\infty} x(-\tau) \xi_- \left( \frac{t - \tau}{\tau} \right) \frac{d\tau}{\tau}. \quad (7.8)$$

For simplicity of exposition, we restrict our subsequent discussion to the case of causal inputs

$$x(t) = 0, \quad t \leq 0$$

and LSI systems whose outputs are causal

$$y(t) = S \{ x(t) \} = 0, \quad t \leq 0.$$

From the development, it will be apparent how to accommodate the more general scenario of (7.8).

For the causal case, only one of the lagged impulse responses (7.6) is required to characterize the system, and, in particular, the input-output relation (7.8) simplifies to

$$y(t) = \int_{0}^{\infty} x(\tau) \xi_+ \left( \frac{t - \tau}{\tau} \right) \frac{d\tau}{\tau} \quad (7.8)$$

where we let $\xi(t) = \xi_+(t)$ to simplify our notation, and where we use the symbol $*$ 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.,

$$x(t) * \xi(t) = \xi(t) * x(t) = \int_{0}^{\infty} x \left( \frac{r}{\tau} \right) \xi(r) \frac{d\tau}{\tau}. \quad (7.10)$$

As a consequence, the cascade of two LSI systems with lagged-impulse responses $\xi_1(t)$ and $\xi_2(t)$, respectively, is typically equivalent to a single system with lagged-impulse response $\xi_1(t) * \xi_2(t)$. Furthermore, such 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.,

$$x(t) * (\xi_1(t) + \xi_2(t)) = x(t) * \xi_1(t) + x(t) * \xi_2(t). \quad (7.11)$$

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

$$x(t) = t^s. \quad (7.12)$$
where $s$ is a complex number. Indeed, from (7.9) and (7.10) the response of an LSI system to (7.12) is readily obtained as

$$y(t) = \hat{\xi}(s) t^s$$

with the associated complex eigenvalue given by

$$\hat{\xi}(s) = \int_0^\infty \xi(t) t^{-s-1} dt$$

whenever this integral converges. Eq. (7.13) is referred to as the Mellin transform of the signal $x(t)$.

The eigenfunction property of the complex power functions implies that the Mellin transform constitutes an important tool in the analysis of LSI systems. 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

$$x(t) = \frac{1}{2\pi j} \int_{c-rej} \hat{X}(s) s \, ds$$

for $t > 0$, where $\hat{X}(s) = \int_0^\infty x(t) t^{-s-1} dt$ and $c$ is in the region of convergence of $\hat{X}(s)$. Eqs. (7.14) collectively constitute the Mellin representation of a signal $x(t)$; the Mellin inverse formula (7.14a) is the synthesis relation, while the Mellin transform (7.14b) 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 transforms $\hat{\xi}(s)$ of their respective lagged-impulse responses. 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 $\hat{\xi}(s)$ can be expressed in the Mellin domain as

$$\hat{Y}(s) = \hat{\xi}(s) \hat{X}(s)$$

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 (7.9) into a convenient multiplicative operation (7.15).

The Mellin transform, its inversion formula, properties, and numerous transform pairs are well documented in the literature. One basic Mellin transform pair is given by

$$t^{-n} u(t - 1) \leftrightarrow \frac{1}{s + s_0}, \quad Re(s) > -s_0$$

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

$$\nabla \hat{x}(t) \equiv \lim_{\varepsilon \to 1} \frac{x(\varepsilon t) - x(t)}{\ln \varepsilon}.$$

One can readily interpret this definition in the context of traditional derivatives as

$$\nabla v_x(t) = \frac{d}{d_{\ln t}} x(t) = t \frac{d}{dt} x(t).$$

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." Consider the following $N$th-order linear constant-coefficient scale-differential equation

$$\sum_{k=0}^N a_k \nabla^k_y y(t) = \sum_{k=0}^M b_k \nabla^k_x x(t),$$

where we denote the $k$th derivative in scale, obtained by iterative application of the derivative operator, by $\nabla^k$. Then, via the convolution property of the Mellin transform, we obtain

$$\hat{Y}(s) = \hat{\xi}(s) \hat{X}(s)$$

where $\hat{\xi}(s)$ is rational, i.e.,

$$\hat{\xi}(s) = k=0^M \frac{b_k s^k}{\prod_{k=0}^N a_k s^k}$$

in the corresponding region of convergence. The usual partial fraction expansion approach, together with Mellin pairs of the form (7.16), can be used to derive $y(t)$ from its Mellin transform.

---

$^1$Actually, we have chosen a slight but inconsequential variant of the Mellin transform—the usual Mellin transform has $s$ replaced by $-s$ in our definition.

$^2$In fact, this development raises some interesting questions regarding connections to the more general literature that is evolving on multiscale systems [84] [108]. Exploring such relationships, however, is beyond the scope of this chapter.
It is interesting to note that in the 1950s, such an approach was developed for the synthesis and analysis of time-varying networks governed by scale-differential and Euler-Cauchy equations, although the relationship to linear scale-invariant system theory was not recognized. Nevertheless, the convolution relationship (7.9) does appear in this work.

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 linear scale-invariant and linear time-invariant (LTI) systems. This relationship between 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 Laplace transform, and linear constant-coefficient scale-differential equations become familiar linear constant-coefficient differential equations.

We next consider a somewhat broader notion of LSI system that will be useful in the sequel.

### 7.3.1 Generalized Linear Scale-Invariant Systems

Suppose $y(t)$ is the response of a system $S\{\cdot\}$ to an arbitrary input $x(t)$. Then we say the system $S\{\cdot\}$ is scale-invariant with parameter $\lambda$ whenever, for any constant $\tau > 0$,

$$S\{x(t/\tau)\} = \tau^{\lambda}y(t/\tau). \quad (7.17)$$

We denote systems that satisfy the superposition principle (7.1) and the generalized scale-invariance relation (7.17) as 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 $\lambda$ is that the kernel satisfy, for any $\alpha > 0$,

$$\kappa(t, \tau) = a^{-(\lambda-1)}\kappa(at, a\tau). \quad (7.18)$$

Such generalized linear scale-invariant systems are also completely characterized in terms of their lagged-impulse response pair (7.6). And, again when we are able to decompose our input according to (7.7) and restrict our attention to the case of causal signals, we can exploit (7.1) and (7.17) to get the following input-output relation

$$y(t) = \int_0^\infty x(\tau) \xi(t/\tau) \frac{d\tau}{\tau^{1-\lambda}}. \quad (7.19)$$

Rewriting (7.19) as

$$y(t) = \int_0^\infty \left(x(\tau)\tau^{\lambda}\right) \xi(t/\tau) \frac{d\tau}{\tau}$$

we observe that in principle any LSI(\lambda) system can be implemented as the cascade of a system that 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.

### 7.4 Linear Time- and Scale-Invariant Systems

We say that a system is linear time- and scale-invariant with parameter $\lambda$, denoted LTI($\lambda$), whenever it jointly satisfies the properties of superposition (7.1), time-invariance (7.2), and generalized scale-invariance (7.17). In this case, the time-invariance constraint (7.3) requires the kernel to be of the form

$$\kappa(t, \tau) = \delta(t - \tau) \quad (7.20)$$

for some impulse response $\delta(\cdot)$, and the scale-invariance constraint (7.18) imposes, in turn, that this impulse response be a generalized homogeneous function of degree $\lambda - 1$, i.e.,

$$\delta(t) = a^{-(\lambda-1)}\delta(at)$$

for all $t$ and all $a > 0$. Following Gel'fand et al. [90], we can parameterize the entire class of impulse responses for such systems. In particular, provided $\lambda \neq 0, -1, -2, \ldots$, we get that $\delta(t)$ takes the form

$$\delta(t) = C_1|t|^{-\lambda}u(t) + C_2|t|^{1-\lambda}u(-t). \quad (7.20a)$$

For the special case $\lambda = -n$ for $n = 0, 1, 2, \ldots$,

$$\delta(t) = C_3|t|^{-(n-1)}u(t) + C_4|t|^{-(n-1)}u(-t) + C_5\delta^{(n)}(t) \quad (7.20b)$$

where $\delta^{(n)}(t)$ denotes the $n$th derivative of the unit impulse and $u(t)$ the unit step function. In both cases, the $C_1, \ldots, C_5$ are arbitrary constants.

There are many familiar LTI($\lambda$) systems. For example, the identity system, for which

$$\delta(t) = \delta(t),$$

corresponds to $\lambda = 0$, $C_3 = C_4 = 0$ and $C_5 = 1$. In fact, as is apparent from the parameterizations (7.20), the identity system is the only stable LTI($\lambda$) system. A second example is the integrator. This system has a regular impulse response

$$\delta(t) = \delta(t),$$

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

$$\delta(t) = \delta(t).$$

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.
7.4.1 Self-Similar Signals and LTSI(\lambda) Systems

In this section, we explore some relationships between self-similar signals and systems. In particular, we show how 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 book.

Our result in the deterministic case is as follows. Let \( v(t) \) be the impulse response of an 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 \( a > 0 \)

\[
y(t) = a^{-(H+\lambda)} y(at).
\]

(7.21)

Two obvious special cases are immediately apparent. The first corresponds to the case in which the system is the identity system (\( \lambda = 0 \)). Here the output and input are identical, and (7.21) yields the appropriate result. The second corresponds to the case in which the input is an impulse (\( H = -1 \)). Here, the output is \( c(t) \), and, again, (7.21) yields the correct result. This, of course, suggests that at least one synthesis for a class of homogeneous signals is in terms of an LTSI(\lambda) system driven by an impulse.

Note that we can derive analogous results for deterministically time-invariant inputs. However, in this case 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 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- 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

\[
y(t) \overset{\text{stat}}{=} a^{-(H+\lambda)} y(at)
\]

with equality in the corresponding statistical sense.
The structure of this kernel imposes certain constraints on the linear system, for example, such systems are symmetric, i.e.,

$$k(t, \tau) = k(\tau, t).$$  

(7.24)

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

$$v^{(b)}_\mu(t-b) = v^{(b)}_\mu+\delta(t)$$

valid for all $b$, the system is time-invariant [i.e., satisfies (7.2)] whenever the multiplier field satisfies

$$K^{(b)}_\nu = K^{(b)}_\nu.$$  

(7.25)

for all $b$. In other words, (7.22) implements an LTI system whenever the field $K^{(b)}_\nu$ is independent of $\nu$. In this case, $K^{(b)}_\nu$ can be expressed as

$$K^{(b)}_\nu = k(\mu)$$

for some regular function of scale $k(\cdot)$.

Likewise, using the identity

$$\zeta^{(a)}_\nu(t) = |a|^{-1/2} \zeta^{(a)}_\nu(\alpha(t))$$

(7.26)

valid for all $a \neq 0$, the system is scale-invariant with parameter $\lambda$ [i.e., satisfies (7.17)] whenever the multiplier field satisfies

$$K^{(a)}_\nu = a^{-\lambda} K^{(a)}_\omega$$

(7.27)

for all $a > 0$.

For the system to be jointly time- and scale-invariant with parameter $\lambda$, (7.25) and (7.27) require that

$$k(\mu) = a^{-\lambda} k(a \mu),$$

i.e., that $k(\cdot)$ be homogeneous of degree $\lambda$. 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 (7.22) is LTSI($\lambda$) whenever the multiplier field has the form

$$K^{(a)}_\nu = C_1 |\mu|^\lambda a(\mu) + C_2 |\mu|^\lambda a(-\mu)$$

(7.28)

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 is even, i.e.,

$$v(t) = v(-t).$$

(7.29)

This is a consequence of the symmetry constraint (7.24). In fact, since we can rewrite (7.23) using (7.26) as

$$h(t, \tau) = \frac{1}{C_1} \int_{-\infty}^{\infty} \mu d\mu \int_{0}^{\infty} \zeta^{(t)}(\mu) \left[ k(\mu) + k(-\mu) \right] \zeta^{(-\tau)}(\mu) \mu^{-\lambda} d\mu.$$

Sec. 7.5  Wavelet-Based LTSI($\lambda$) Systems

(7.30)

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 (7.28) and choose

$$K^{(a)}_\nu = k(\mu) = C \mu^\lambda a(\mu)$$

(7.30)

where $C$ is an arbitrary constant.

Finally, combining (7.29) with (7.30), we can conclude that whenever (7.22) implements an LTSI($\lambda$) system, i.e., whenever $k(\cdot)$ is chosen according to (7.30), the impulse response corresponding to the system of (7.22) must take the form

$$v(t) = C |t|^\lambda$$

for $\lambda \neq 0, -1, -2, -3, \ldots$, or the form

$$v(t) = C |t|^{-n-1} + C_n \delta(n)$$

for $\lambda = n$ for $n = 0, 2, 4, \ldots$. In both cases, $C$, and $C_n$ are parameters determined by the constant $C$ in (7.30). Note in particular that, at least for the case $\lambda = 0$, we must have

$$C_n = 0.$$

This follows from the fact that, since $K^{(a)}_\nu \equiv C$, the overall system (7.22) 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 $\lambda$. 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 (7.29) precludes us from being able to implement either the differentiator or integrator system examples discussed in Section 7.4 since these systems have impulse responses that are not even.

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 development, it is reasonable to expect it to be a factor in any practical implementation. In the next section, we consider a strategy for approximating LTSI($\lambda$) systems that exploits orthonormal wavelet bases. As we will see, these quasi-LTSI($\lambda$) systems are particularly convenient to implement and can be made computationally efficient.

7.5.1 Dyadic Approximations to LTSI($\lambda$) Systems

A practical approximation to a linear time-scale invariant system can be constructed via orthonormal wavelet bases of the type described in Section 2.3.
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 LTSI(X) 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.,

\[
x^m_n = \int_{-\infty}^{\infty} x(t) \psi^n_m(t) \, dt,
\]

then scales the resulting collection of wavelet coefficients by a factor \( k^m_n \),

\[
y^m_n = k^m_n x^m_n
\]

then resynthesizes a signal from these modified coefficients to generate an output according to the synthesis formula (2.5a), i.e.,

\[
y(t) = \sum_m \sum_n y^m_n \psi^n_m(t).
\]

It is a straightforward exercise to show that the overall system, described via

\[
y(t) = W_d^{-1} \{ k^m_n W_d \{ x(t) \} \},
\]

(7.31)

corresponds to a symmetric linear system with kernel

\[
\tilde{k}(t, \tau) = \sum_m \sum_n \psi^n_m(t) k^m_n \psi^n_m(\tau).
\]

A close inspection reveals that, as a consequence of the nature of the discretization inherent in the system, one cannot choose the multiplier coefficients \( k^m_n \) 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 \( \lambda \). However, we can show that if the \( k^m_n \) are chosen in a manner consistent with the discussion of the previous section, viz.,

\[
k^m_n = C \mu \big|_{\mu=2^{-m}} = C 2^{-\lambda m},
\]

(7.32)

then the system defined via (7.31) 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, for any \( m \),

\[
\tilde{k}(t, \tau) = 2^{-(\lambda-1)m} \tilde{k}(2^m t, 2^m \tau),
\]

(7.33)

where we have used the identity

\[
\psi^n_m(2t) = 2^{-i/2} \psi^{m+n}_m(t),
\]

valid for any integer \( i \). However, since (7.33) can be restated in terms of the generalized scale invariance condition (7.18) as

\[
\tilde{k}(t, \tau) = a^{-\lambda-1} \tilde{k}(a \tau, a t), \quad a = 2^m
\]

we see that the system obeys a weaker, dyadic scale invariance condition. In particular, the system satisfies (7.17) only for dilation factors \( \tau \) 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 that have no detail at scales coarser than \( 2^M \) for some integer \( M \), so that

\[
x^m_n = 0, \quad m < M.
\]

In this case, the multiplier coefficients \( k^m_n \) for \( m < M \) for the system are irrelevant and we may arbitrarily assume them to be zero. For this class of inputs, the effective kernel is

\[
\tilde{k}_{eff}(t, \tau) = \sum_{m \geq M} \sum_n \psi^n_m(t) C 2^{-\lambda m} \psi^n_m(\tau).
\]
Using the identity 
\[ \psi_m^n(t - 2^l - M) = \psi_m^n(t - 2^l - M) \]
valid for \( m \geq M \) and \( l \) an integer, we see that this kernel satisfies 
\[ \tilde{\eta}(t, \tau) = \tilde{\eta}(t - 12^{-M} - \tau - 12^{-M}) \]
for all integers \( l \). Since (7.34) can be re-expressed as 
\[ \tilde{\eta}(t, \tau) = \tilde{\eta}(t - h, \tau - h), \quad h = 2^{-M} \]
we see that for this class of input signals the system is periodically time-varying, i.e., satisfies (7.2) for any shift factor \( \tau \) of the form 
\[ \tau = 2^{-M}, \quad l = \ldots, -1, 0, 1, 2, \ldots \]

Note that in contrast to the wavelet-based systems discussed in the previous section, in this case the actual choice of wavelet affects the characteristics of the overall system. Indeed, with respect to scaling behavior, the choice of wavelet affects how the system behaves under nondyadic scale changes at the input. Furthermore, the choice of wavelet affects 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 \( 2^{-M} \).

### 7.6 SUMMARY

In this chapter, we undertook a preliminary investigation into the system theoretic foundations of the concepts developed in this book. This was aimed toward developing some unifying perspectives on the results we have obtained. After defining scale-invariant systems, we explored the relationships between such systems, self-similar signals, and the wavelet transform. Our results provide additional evidence that wavelet-based synthesis, analysis, and processing of self-similar signals are rather natural. Indeed, the 1/f synthesis and whitening filters described in Section 4.2—which play an important role in detection and estimation problems involving 1/f processes—are specific examples of linear systems that are effectively jointly time- and scale-invariant. Interpreting the transmitter and receiver structures for fractal modulation discussed in Chapter 6 in terms of such systems also has the potential to 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, our results suggest 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 jointly time- and scale-invariant. Moreover, we showed that wavelet representations lead to some very efficient and practical computational structures for characterizing and implementing such systems. Ultimately, the ideas developed in this chapter may well lead to a basis for a unified development of fractal signal and system theory. As such, this represents one of several interesting and potentially rich open directions for further research.
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

\[ \phi_0^n(t), \psi_0^n(t) \in V_0 \subset V_1, \]

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.,

\[ \phi_0^n(t) = \sum_i h[i] \phi_i(t) \]  
\[ \psi_0^n(t) = \sum_i g[i] \phi_i(t) \]

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

\[ \text{In any case, multiplying both sides of (A.1)} \]

where, in turn, we may rewrite (2.20) as

\[ h[l - 2n] = \int_{-\infty}^{\infty} \phi_0^m(t) \phi_0^{m+1}(t) \, dt, \]  
\[ g[l - 2n] = \int_{-\infty}^{\infty} \psi_0^m(t) \phi_0^{m+1}(t) \, dt. \]

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

we can write

\[ \phi_0^{m+1}(t) = A_m \{ \phi_0^n(t) \} + D_m \{ \phi_0^n(t) \} \]

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

\[ \phi_0^{m+1}(t) = \int_{-\infty}^{\infty} x(t) \phi_0^{m+1}(t) \, dt. \]
Appendix B

Proofs for Chapter 3

B.1 Proof of Theorem 3.2

Let \( \omega_0 \) and \( \omega_1 \) be constants from Definition 3.1, and let \( \lambda = \omega_1/\omega_0 \). 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

\[
B_\lambda(\omega) = \begin{cases} 
1 & a\omega_0 < |\omega| \leq a\omega_1 \\
0 & \text{otherwise} 
\end{cases} 
\]

for any \( a > 0 \), the output \( y_\lambda(t) \) is wide-sense stationary, has finite variance, and has an autocorrelation satisfying

\[
R_{y_\lambda}(\tau) = E[y_\lambda(t) y_\lambda(t - \tau)] = a^{-2H} R_{y_\lambda}(a\tau) 
\]

for all \( a > 0 \). Furthermore, for any distinct integers \( m \) and \( k \), the processes \( y_{M^k}(t) \) and \( y_{M^m}(t) \) are jointly wide-sense stationary.

**Proof:**

First, from Definition 3.1 we have immediately that \( y_\lambda(t) \) is wide-sense stationary. More generally, consider the case \( a > 0 \). Let \( b_\lambda(t) \) be the impulse response of the filter with frequency response (B.1). To establish (B.2), it suffices to note that \( y_\lambda(t) \) has correlation function

\[
R_{y_\lambda}(t, s) = E[y_\lambda(t) y_\lambda(s)]
\]

where we have exploited the identities (3.2b) and

\[
b_\lambda(t) = ab_\lambda(at).
\]

However, since \( y_\lambda(t) \) is wide-sense stationary, the right side of (B.3) is a function only of \( t - s \). Hence, \( y_\lambda(t) \) is wide-sense stationary and (B.2) follows. Furthermore, \( y_\lambda(t) \) has variance

\[
R_{y_\lambda}(0, 0) = a^{-2H} R_{y_\lambda}(0, 0) < \infty
\]

where the inequality is a consequence of Definition 3.1. To establish our final result, since \( B_{M^k}(\omega) \) and \( B_{M^m}(\omega) \) occupy disjoint frequency intervals for \( m \neq k \), the spectra of \( y_{M^k}(t) \) and \( y_{M^m}(t) \) likewise occupy disjoint frequency intervals. Thus, \( y_{M^k}(t) \) and \( y_{M^m}(t) \) are uncorrelated, and, hence, jointly wide-sense stationary as well.

Proceeding now to a proof of our main theorem, let us establish that \( y(t) \) is wide-sense stationary. Let \( M_L \) and \( M_U \) be any pair of integers such that

\[
\lambda^{M_U} \omega_0 < \omega_L < \omega_U < \lambda^{M_L} \omega_1
\]

and consider preceding the filter (3.25) with a filter whose frequency response is

\[
\tilde{B}(\omega) = \begin{cases} 
1 & \lambda^{M_L} \omega_0 < |\omega| \leq \lambda^{M_U} \omega_1 \\
0 & \text{otherwise} 
\end{cases}
\]

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

\[
\tilde{B}(\omega) = \sum_{m=M_L}^{M_U} B_{\lambda^m}(\omega)
\]

where \( B_{\lambda^m}(\omega) \) is as defined in (B.1) of Lemma B.1, we can decompose \( \tilde{y}(t) \) according to

\[
\tilde{y}(t) = \sum_{m=M_L}^{M_U} y_{\lambda^m}(t)
\]

where \( y_{\lambda^m}(t) \) is the response of the filter with frequency response \( B_{\lambda^m}(\omega) \) to \( x(t) \). Since, by Lemma B.1, all the terms comprising the summation (B.5) are jointly wide-sense stationary, \( \tilde{y}(t) \) 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 \( \tilde{y}(t) \) [40].
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
\[
S_{y_p}(\omega) = a^{-2(2H+1)}S_{y_p}(\omega)
\]
where \( S_{y_p}(\omega) \) is the power spectrum associated with \( y_p(t) \). For \( 1 < a < \lambda \), we observe that \( S_{y_p}(\omega) \) and \( S_{y_p}(\omega) \) have spectral overlap in the frequency range \( \omega_0 < |\omega| < \omega_1 \), and we can therefore conclude that the two spectra must be identical in this range. The reasoning is as follows. If we pass either \( y_p(t) \) or \( y_1(t) \) through the bandpass filter with frequency response whose impulse response is \( b(t) \), the outputs must identical, i.e.,
\[
y_p(t) * y_p(t) = y_1(t) * y_1(t) = y_1(t) * x(t).
\]
Since \( y_p(t) \) and \( y_1(t) \) are jointly wide-sense stationary, we then conclude
\[
S_{y_p}(\omega) = S_{y_1}(\omega), \quad \omega_0 < |\omega| < \omega_1.
\]
whence
\[
S_{y_p}(\omega) = S_{y_1}(\omega), \quad a \omega_0 < |\omega| < \omega_1.
\]
Combining (B.7) with (B.6) we get
\[
S_{y_p}(\omega) = a^{-2(2H+1)}S_{y_p}(\omega), \quad a \omega_0 < |\omega| < \omega_1
\]
for any \( 1 < a < \lambda \). Differentiating (B.8) with respect to \( a \) and letting \( a \to 1+ \), we find that
\[
\omega S_{y_p}'(\omega) = -(2H+1)S_{y_p}(\omega), \quad a \omega_0 < |\omega| < \omega_1
\]
and note that all positive, even, regular solutions to this equation are of the form
\[
S_{y_p}(\omega) = \sigma_z^2|\omega|^{\gamma}, \quad \omega_0 < |\omega| < \omega_1
\]
for some \( \sigma_z^2 > 0 \) and \( \gamma = 2H + 1 \). Using (B.9) with (B.6) we find, further, that
\[
S_{y_1}(\omega) = \sigma_z^2|\omega|^{\gamma} \lambda^M \omega_0 < |\omega| \leq \lambda^M \omega_1
\]
where \( H' = \gamma - 1 \). Via Lemma B.1, the \( y_1(t) \) are uncorrelated, so we deduce that \( y(t) \) has spectrum
\[
S_y(\omega) = \sum_{m=M}^{M+2} S_{y_1}(\omega) = \left\{ \begin{array}{ll}
\sigma_z^2|\omega|^{\gamma} \lambda^M \omega_0 < |\omega| \leq \lambda^M \omega_1 \\
0 & \text{otherwise}
\end{array} \right.
\]
Finally, since
\[
S_y(\omega) = |B(\omega)|^2 S_y(\omega)
\]
our desired result (3.26) follows.
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

\[
x(t) = \lim_{M \to \infty} x_M(t) = \sum_m \sum_n x_m^n \psi_n^m(t).
\]

Since for each \( m \) the wavelet coefficient sequence \( x_m^n \) is wide-sense stationary with spectrum \( 2^{-m} \), the approximation \( x_M(t) \) is cyclostationary [40] with period \( 2^M \), has finite variance, and has the associated time-averaged spectrum

\[
S_M(\omega) = \sum_{m=-M}^M 2^{-m} |\Psi(2^{-m} \omega)|^2.
\]  

(B.12)

The limiting time-averaged spectrum

\[
S_k(\omega) = \lim_{M \to \infty} S_M(\omega)
\]

gives the desired spectrum expression (3.36), and corresponds to the time-averaged spectrum of \( \pi(t) \) as measured at the output of a bandpass filter for each frequency \( \omega \) 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 \( \omega \), we can choose \( m_0 \) and \( \omega_0 \) such that \( \omega = 2^{m_0} \omega_0 \) and \( 1 \leq |\omega_0| < 2 \). Hence, using (3.36) we see

\[
S_k(\omega) = 2^{-m_0} S_k(\omega_0)
\]

from which it follows that

\[
\left[ \inf_{|\omega| < 2} S_k(\omega_0) \right] \frac{1}{|\omega|} \leq S_k(\omega) \leq \left[ \sup_{|\omega| < 2} S_k(\omega_0) \right] \frac{2^\pi}{|\omega|}.
\]

It suffices, therefore, to find upper and lower bounds for \( S_k(\omega_0) \) 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 \( \omega \to \infty \). 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

\[
S_k(\omega_0) \leq \sum_{m=0}^\infty 2^{-m} + \sum_{m=1}^\infty 2^{-m} C^2 2^{-2Rm} < \infty.
\]

To establish the lower bound it suffices to show \( S_k(\omega) > 0 \) for every \( 1 \leq \omega \leq 2 \), which we establish by contradiction.

Suppose for some \( 1 \leq \omega_0 \leq 2 \),

\[
S_k(\omega_0) = \sum_m 2^{-m} |\Psi(2^{-m} \omega_0)|^2 = 0
\]

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 \( S_0(\omega) > 0 \) for every \( 1 \leq \omega_0 \leq 2 \). The complete theorem follows.

8.4 PROOF OF THEOREM 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

\[
x_K(t) = \lim_{K \to \infty} x_K(t) = x(t).
\]

Then by Theorem 3.2, \( x_K(t) \) is wide-sense stationary and has power spectrum

\[
S_K(\omega) = \begin{cases} 
\sigma_0^2/|\omega| & 2^{-K} < |\omega| \leq 2^K \\
0 & \text{otherwise}
\end{cases}
\]  

(B.13)

If we denote its corresponding autocorrelation by

\[
R_K(\tau) = E[x_K(t)x_K(t+\tau)]
\]

and its wavelet coefficients by

\[
x_m^K(t) = \int_{-\infty}^{\infty} x_K(t) \psi_m^k(t) \, dt,
\]

the correlation between wavelet coefficients may be expressed as

\[
E[x_m^K(\tau)x_n^K(\tau)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi_m^k(t) R_K(t-\tau) \psi_n^k(\tau) \, dt \, d\tau
\]

(B.14)

Applying Parseval's theorem and exploiting (B.13), we may rewrite (B.14) in the frequency domain as

\[
E[x_m^K(\tau)x_n^K(\tau)] = \frac{2^{-(m+n+1)/2}}{2\pi^{1/R}} \int_{-2\pi}^{2\pi} \sqrt{2^{2R}} \frac{\sigma_0^2}{|\omega|} \Psi(2^{-m} \omega) \Psi(2^{-n} \omega) \, d\omega
\]

\[
+ \int_{2\pi}^{2\pi} \frac{\sigma_0^2}{|\omega|} \Psi(2^{-m} \omega) \Psi(2^{-n} \omega) \, d\omega.
\]  

(B.15)
Interchanging limits, we get
\[ x_n^m = \lim_{K \to \infty} x_n^m(K) \]
and, in turn,
\[ E \left[ x_n^m x_{n'}^m \right] = \lim_{K \to \infty} E \left[ x_n^m(K) x_{n'}^m(K) \right]. \]  
Substituting (B.15) into (B.16) yields (3.40). Since
\[ |E \left[ x_n^m x_{n'}^m \right]| \leq \text{var} x_n^m \cdot \text{var} x_{n'}^m, \]
and since
\[ \text{var} x_n^m = \frac{\sigma_r^2 2^{-nm}}{\pi} J \]
where
\[ J = \int_0^\infty \omega^{-\gamma} |\Psi(\omega)|^2 d\omega, \]
it suffices to show that (B.17) converges. Because \( \psi(t) \) has \( R \) vanishing moments, there exist constants \( C_0 \) and \( C_1 \) such that
\[ |\Psi(\omega)| < C_0 |\omega|^R, \]
and since
\[ |\Psi(\omega)| < C_1 |\omega|^{-R}. \]
Using (B.18) in (B.17), we obtain, for \( 0 < \gamma < 2R \) and \( R \geq 1 \),
\[ J = \int_0^1 C_2^2 \omega^{-2R-\gamma} d\omega + \int_1^\infty C_2^2 \omega^{-2R-\gamma} d\omega < \infty \]
as required.

**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
\[ R_{m_n m'_n} = \frac{\sigma_r^2}{\pi \sigma_{\sigma^2}} \text{Re} I(\Delta) \]
where
\[ I(\Delta) = \int_0^\infty \Xi(\omega) e^{-i\Delta \omega} d\omega. \]
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 \( \Delta \). Indeed, since
\[ \Xi(\omega) \sim O\left(\omega^{2R-\gamma}\right), \quad \omega \to 0 \]
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 [50], that
\[ I(\Delta) \to 0, \quad \Delta \to \infty. \]
When \( 0 < \gamma < 2R \), we may integrate (B.20) by parts \( Q \) times, for some positive integer \( Q \), to obtain
\[ I(\Delta) = \frac{1}{(j+1)\Delta^j} \int_0^\infty \Xi(Q)(\omega) e^{-i\Delta \omega} d\omega \]
\[ + \sum_{q=0}^{Q-1} \frac{1}{(j+1)\Delta^j} \left\{ \lim_{\omega \to 0} \Xi^{(q)}(\omega) e^{-i\Delta \omega} - \lim_{\omega \to \infty} \Xi^{(q)}(\omega) e^{-i\Delta \omega} \right\}. \]
Due to the vanishing moments of the wavelet we have
\[ \Xi^{(q)}(\omega) \sim O\left(\omega^{2R-\gamma-q}\right), \quad \omega \to 0 \]
while due to the regularity of the wavelet, \( \Psi(\omega) \) decays at least as fast as \( 1/\omega^R \) as \( \omega \to \infty \), whence
\[ \Xi^{(q)}(\omega) \sim O\left(\omega^{2R-\gamma-q}\right), \quad \omega \to \infty. \]
Hence, the limit terms in (B.21) for which \( -2R - \gamma + 1 < q < 2R - \gamma \) all vanish.
Moreover, when we substitute \( q = Q \), (B.22) and (B.23) imply that \( \Xi^{(Q)}(\omega) \) is absolutely integrable, i.e.,
\[ \int_0^\infty |\Xi^{(Q)}(\omega)| d\omega < \infty, \]
whenever \( -2R - \gamma + 1 < Q < 2R - \gamma + 1 \), which implies, again via the Riemann-Lebesgue lemma, that the integral in (B.21) vanishes asymptotically, i.e.,
\[ \int_0^\infty \Xi^{(Q)}(\omega) e^{-i\Delta \omega} d\omega \to 0, \quad \Delta \to \infty. \]
Hence, choosing \( Q = [2R - \gamma] \) in (B.21) (so \( 2R - \gamma \leq Q < 2R - \gamma + 1 \)) allows us to conclude
\[ I \sim O\left(\Delta^{-[2R-\gamma]}\right), \quad \Delta \to \infty. \]
Substituting (B.26) into (B.19) 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 \( \Theta = \{\beta, \sigma^2, \sigma^2_w\} \) for the scenario described in Section 4.3.

We begin by defining our observed (incomplete) data to be
\[
 r = \{r_m, m, n \in \mathbb{R}\},
\]
and our complete data to be \((x, r)\) where
\[
 x = \{x^m_n, m, n \in \mathbb{R}\}.
\]
Consequently, the EM algorithm for the problem is defined as [80]

E step: Compute
\[
 U(\Theta, \hat{\Theta}^{(i)}).
\]

M step:
\[
 \max_{\Theta} U(\Theta, \hat{\Theta}^{(i)}) \rightarrow \hat{\Theta}^{(i+1)}
\]
where
\[
 U(\Theta, \hat{\Theta}) \triangleq E \left[ \ln p_{x, r}(x, r; \Theta) \right].
\]

For our case, \( U \) is obtained conveniently via
\[
 U(\Theta, \hat{\Theta}) = E \left[ \ln p_{x, r}(x, r; \Theta) + \ln p_{x}(x; \Theta) \right] r; \hat{\Theta}
\]
with
\[
 p_{x, r}(x; \Theta) = \prod_{m, n \in \mathbb{R}} \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left\{ -\frac{(x^m_n - r_m)^2}{2\sigma^2} \right\}
\]

and
\[
 p_{x}(x; \Theta) = \prod_{m, n \in \mathbb{R}} \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left\{ -\frac{(x^m_n)^2}{2\sigma^2} \right\}.
\]

Then
\[
 U(\Theta, \hat{\Theta}) = -\frac{1}{2} \sum_{m, n \in \mathbb{R}} N(m) \left\{ \frac{1}{\sigma^2_w} S^m_m(\Theta) + \ln 2\pi \sigma^2_w + \frac{1}{\sigma^2 - \sigma^2_m} S^m_m(\Theta) + \ln 2\pi \sigma^2 \right\}
\]

where
\[
 S^m_m(\Theta) = \frac{1}{N(m)} \sum_{n \in X(m)} E \left[ (x^m_n)^2 | r_m, \Theta \right]
\]
\[
 S^m_m(\Theta) = \frac{1}{N(m)} \sum_{n \in X(m)} E \left[ (x^m_n)^2 | r_m, \Theta \right]
\]
are (quasi) conditional sample-variance estimates from the data based upon the model parameters \( \Theta \). Evaluating the expectations we get
\[
 S^m_m(\Theta) = A_m(\Theta) + B^m_m(\Theta) \sigma^2_w
\]
\[
 S^m_m(\Theta) = A_m(\Theta) + B^m_m(\Theta) \sigma^2_w
\]

where
\[
 A_m(\Theta) = \frac{\sigma^2_w \cdot \sigma^2 \sigma^2 - \sigma^2 \sigma^2}{\sigma^2_w + \sigma^2 \sigma^2 - \sigma^2}
\]
\[
 B^m_m(\Theta) = \frac{\sigma^2_w \cdot \sigma^2 \sigma^2 - \sigma^2 \sigma^2}{\sigma^2_w + \sigma^2 \sigma^2 - \sigma^2}
\]

which completes our derivation of the E step.

To derive the structure of the M step, we maximize \( U(\Theta, \hat{\Theta}) \) as given by (C.1). This maximization is always well defined as \( U(\Theta, \Theta) \leq L(\Theta) \) for any \( \Theta, \Theta \).

The local extrema are obtained by differentiating \( U(\Theta, \hat{\Theta}) \) with respect to each of the parameters of \( \Theta \). Since (C.1) expresses \( U(\Theta, \hat{\Theta}) \) as the sum of two terms, one of which depends only on \( \sigma^2_w \) and the other of which depends only on \( \beta \) and \( \sigma^2 \), the maximization can be broken down into two independent parts.

\[
 p_{x, r}(x, r; \Theta) = \prod_{m, n \in \mathbb{R}} \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left\{ -\frac{(x^m_n - r_m)^2}{2\sigma^2} \right\}
\]
Considering first our maximization over $\sigma^2_{uv}$, we readily obtain the maximizing $\sigma^2_{uv}$ as the sample-average

$$\hat{\sigma}^2_{uv} = \frac{\sum_{m \in M} N(m)S_m^u(\hat{\Theta})}{\sum_{m \in M} N(m)}.$$

Turning next to $\beta$ and $\sigma^2$, we find that the maximizing parameters $\beta$ and $\sigma^2$ satisfy

$$\sum_{m \in M} N(m)S_m^u(\hat{\Theta})\beta^m = \sigma^2 \sum_{m \in M} N(m), \quad (C.2a)$$
$$\sum_{m \in M} mN(m)S_m^u(\hat{\Theta})\beta^m = \sigma^2 \sum_{m \in M} mN(m). \quad (C.2b)$$

Eliminating $\sigma^2$ we obtain that $\beta$ is the solution of the polynomial equation

$$\sum_{m \in M} C_m N(m)S_m^u(\hat{\Theta}) \beta^m = 0, \quad (C.3)$$

where $C_m$ is as defined in (4.16). The eliminated variable $\sigma^2$ is trivially obtained by back-substitution:

$$\hat{\sigma}^2 = \frac{\sum_{m \in M} N(m)S_m^u(\hat{\Theta})\beta^m}{\sum_{m \in M} N(m)}.$$

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

$$\sum_{m \in M} C_m K_m \beta^m = 0 \quad (C.4)$$

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

$$m_* = \frac{\sum_{m \in M} mN(m)}{\sum_{m \in M} N(m)}$$

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, $C_m(m_m - m_*) \geq 0$ 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

$$f'(\beta) \equiv f(\beta)\beta^{-m}.$$ 

is increasing for $\beta > 0$, i.e.,

$$f'(\beta) = \sum_{m \in M} C_m(m - m_*)N(m)\beta^{m-m_*} > 0.$$

Then, since $f(0) = -\infty$ and $f(\infty) = \infty$, we see $f(\beta)$ 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

To show that $y(t)$ has finite energy, we exploit an equivalent synthesis for $y(t)$ as the output of a cascade of filters driven by $x(t)$, the first of which is an ideal bandpass filter whose passband includes $\omega_L < |\omega| < \omega_U$, and the second of which is the filter given by (5.4).

Let $b_m(t)$ be the impulse response of a filter whose frequency response is given by

$$B_m(\omega) = \begin{cases} 1 & 2^{-m} \pi < |\omega| \leq 2^{m+1} \pi \\ 0 & \text{otherwise} \end{cases} \quad (D.1)$$

and let $b(t)$ be the impulse response corresponding to (5.4). Furthermore, 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, using $\ast$ to denote convolution,

$$y(t) = b(t) \ast \left[ \sum_{m=M_L}^{M_U} b_m(t) \right] \ast x(t)$$

$$= b(t) \ast \left[ \sum_{m=M_L}^{M_U} \tilde{x}_m(t) \right] \quad (D.2)$$

where

$$\tilde{x}_m(t) = x(t) \ast b_m(t) = 2^{-mH} \hat{x}_0(2^m t). \quad (D.3)$$

and where the last equality in (D.3) results from an application of the self-similarity relation (5.2) and the identity

$$b_m(t) = 2^m b_0(2^m t).$$

Because $x(t)$ is energy-dominated, $\hat{x}_0(t)$ has finite energy. Hence, (D.3) implies that every $\tilde{x}_m(t)$ has finite energy. Exploiting this fact in (D.2) allows us to conclude that $y(t)$ must have finite energy as well.

To verify the spectrum relation (5.5), we express (D.2) in the Fourier domain. Exploiting the fact that we may arbitrarily extend the limits in the summation in (D.2), we get

$$Y(\omega) = B(\omega) \sum_{m=-\infty}^{\infty} \hat{X}_m(\omega) = \begin{cases} X(\omega) & \omega_L < |\omega| < \omega_U \\ 0 & \text{otherwise} \end{cases}$$

where $\hat{X}_m(\omega)$ denotes the Fourier transform of $\tilde{x}_m(t)$, and where

$$X(\omega) \triangleq \sum_{m=-\infty}^{\infty} \hat{X}_m(\omega). \quad (D.4)$$

The right-hand side of (D.4) is, of course, pointwise convergent because for each $\omega$ at most one term in the sum is non-zero. Finally, exploiting (D.3) in (D.4) gives

$$X(\omega) = \sum_{m} 2^{-m(H+1)} \hat{x}_0(2^{-m} \omega),$$

which, as one can readily verify, satisfies (5.6).

D.2 PROOF OF THEOREM 5.3

To prove the "only if" statement, we suppose $x(t) \in \mathcal{B}^H$ and begin by expressing $x(t)$ in terms of the ideal bandpass wavelet basis. In particular, we let

$$x(t) = \sum_{m} \tilde{x}_m(t)$$

where

$$\tilde{x}_m(t) = \beta^{-m/2} \sum_n q[n] \tilde{q}_m^* n(t)$$

and where $\tilde{q}[n]$, the generating sequence in this basis, has energy $\tilde{E} < \infty$. The new generating sequence $q[n]$ can then be expressed as

$$q[n] = \sum_m q_m[n] \quad (D.5)$$

where

$$q_m[n] = y_m(t)|_{t=n}$$
and

\[ y_m(t) = \hat{x}_m(t) * \psi(t). \]

For each \( m \), since \( \hat{x}_m(t) \) is bandlimited, \( y_m(t) \) and \( q_m[n] \) each have finite energy and Fourier transforms \( Y_m(\omega) \) and \( Q_m(\omega) \) respectively. Hence,

\[ Q_m(\omega) = \sum_k Y_m(\omega - 2\pi k) \tag{D.6} \]

where

\[ Y_m(\omega) = \begin{cases} (2\beta)^{-m/2} \psi^*(\omega) \hat{Q}(2^{-m}\omega) 2^{-m} \pi < |\omega| \leq 2^{m+1}\pi \\ 0 \quad \text{otherwise} \end{cases} \]

with \( \hat{Q}(\omega) \) denoting the Fourier transform of \( \hat{q}(n) \), and \( \Psi^*(\omega) \) the complex conjugate of \( \Psi(\omega) \).

In deriving bounds on the energy \( E_m \) in each sequence \( q_m[n] \) for a fixed \( m \), it is convenient to consider the cases \( m \leq -1 \) and \( m \geq 0 \) separately. When \( m \leq -1 \), the sampling by which \( q_m[n] \) is obtained involves no aliasing. Since on \( [-\pi, \pi) \) we then have

\[ Q_m(\omega) = Y_m(\omega). \]

we may deduce that \( q_m[n] \) has energy

\[ E_m = \sum_n |q_m[n]|^2 = \frac{(2\beta)^{-m}}{\pi} \int_{2^{-m}\pi}^{2^{m+1}\pi} |\Psi(\omega)|^2 |\hat{Q}(2^{-m}\omega)|^2 d\omega. \tag{D.7} \]

Because \( \psi(t) \) has \( R \) vanishing moments, there exists a \( 0 < \epsilon_0 < \infty \) such that

\[ |\Psi(\omega)| \leq \epsilon_0 |\omega|^{-R} \tag{D.8} \]

for all \( \omega \). Exploiting this in (D.7) we obtain

\[ E_m \leq C_1 2^{(2R-\gamma)m} \tilde{E} \tag{D.9} \]

for some \( 0 \leq C_0 < \infty \).

Consider, next, the case corresponding to \( m \geq 0 \). Since \( \psi(t) \) has \( R \) vanishing moments, there also exists a \( 0 < \epsilon_1 < \infty \) such that

\[ |\Psi(\omega)| \leq \epsilon_1 |\omega|^{-R} \tag{D.10} \]

for all \( \omega \). Hence, on \( 2^m \pi < |\omega| \leq 2^{m+1}\pi \),

\[ |Y_m(\omega)| \leq \epsilon_1 \pi^{-R} 2^{-(\gamma+1+2R)m/2} |\hat{Q}(2^{-m}\omega)|. \tag{D.11} \]

From (D.6), we obtain

\[ |Q_m(\omega)| \leq \epsilon_1 \pi^{-R} 2^{-(\gamma+1+2R)m/2} \sum_{k=0}^{2^m-1} |\hat{Q}(2^{-m}\omega + 2\pi k 2^{-m})| \tag{D.12} \]

by exploiting, in order, the triangle inequality, the bound (D.11), the fact that only \( 2^m \) terms in the summation in (D.6) are non-zero since \( y_m(t) \) is bandlimited, and the fact that \( \hat{Q}(\omega) \) is \( 2\pi \)-periodic. In turn, we may use, in order, (D.12), the Schwarz inequality, and again the periodicity of \( \hat{Q}(\omega) \) to conclude that

\[ E_m \leq \epsilon_1^2 \pi^{-2R} 2^{-(\gamma+1+2R)m} \left[ \sum_{k=0}^{2^m-1} \frac{1}{2\pi} \int_{2\pi}^{2^m \pi} |\hat{Q}(2^{-m}\omega + 2\pi k 2^{-m})|^2 d\omega \right]^2 \leq C_2 2^{-(\gamma+2+4R)m} E \tag{D.13} \]

for some \( 0 \leq C_1 < \infty \).

Using (D.5), the triangle inequality, and the Schwarz inequality, we obtain the following bound on the energy in \( q[n] \)

\[ E = \sum_n |q[n]|^2 \leq \left[ \sum_m \sqrt{E_m} \right]^2 \]

which from (D.13) and (D.9) is finite provided \( 0 < \gamma < 2R \) and \( R \geq 1 \).

Let us now show the converse. Suppose \( q[n] \) has energy \( E < \infty \), and express \( x(t) \) as

\[ x(t) = \sum_m x_m(t) \]

where

\[ x_m(t) = \beta^{-m/2} \sum_n q[n] \psi^m_n(t). \]

If we let

\[ \tilde{y}_m(t) = b_0(t) * x_m(t) \]

where \( b_0(t) \) is the impulse response of the ideal bandpass filter in Definition 5.1, it suffices to show that

\[ \tilde{y}(t) = \sum_m \tilde{y}_m(t) \tag{D.14} \]

has finite energy.

For each \( m \), we begin by bounding the energy in \( \tilde{y}_m(t) \), which is finite because \( x_m(t) \) has finite energy. Since \( \tilde{y}_m(t) \) has Fourier transform

\[ \tilde{Y}_m(\omega) = \begin{cases} (2\beta)^{-m/2} \hat{Q}(2^{-m}\omega) \Psi(2^{-m}\omega) & \pi |\omega| \leq 2\pi \\ 0 & \text{otherwise} \end{cases} \]

where \( \hat{Q}(\omega) \) is the discrete-time Fourier transform of \( q[n] \), we get

\[ \tilde{E}_m = \frac{2^{-m}}{\pi} \int_{2^{-m}\pi}^{2^{m+1}\pi} |\hat{Q}(\omega)|^2 |\Psi(2^{-m}\omega)|^2 d\omega. \]

Again, it is convenient to consider the cases corresponding to \( m \leq -1 \) and \( m \geq 0 \) separately. For \( m \leq -1 \), most of the energy in \( x_m(t) \) is at frequencies
Proofs for Chapter 5

App. D

Sec. D.4 Proof of Theorem 5.6

We first establish some notation. Let us denote the cross-correlation between two finite-power signals \( f(t) \) and \( g(t) \) by

\[
R_{fg}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f(t) g(t - \tau) \, dt.
\]

Its Fourier transform is the corresponding cross-spectrum \( S_{fg}(\omega) \). Similarly

\[
R_{ab}[k] = \lim_{L \to \infty} \frac{1}{2L + 1} \sum_{n=-L}^{L} a[n] b[n - k]
\]

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

\[
x_{m}(t) = \beta^{-m/2} \sum_{n} q[n] \tilde{y}(t - 2^{-m}n)
\]

and note that

\[
\tilde{y}(t) = \sum_{n} q[n] \delta(t - 2^{-m}n)
\]

where \( q[n] \) is an upsampled version of \( q[n] \), i.e.,

\[
q[n] = \begin{cases} 
q[2^{-m}n] & n = 2^{ml}, \; l = \ldots, -1, 0, 1, 2, \ldots \\
0 & \text{otherwise}
\end{cases}
\]

and note that

\[
v_{m}(t) = \beta^{-m/2} \sum_{n} q[n] \delta(t - 2^{-m}n)
\]

which, as one can readily verify, satisfies (5.24).
166

Proofs for Chapter 5

where

\[ R_q[k] = \lim_{L \to \infty} \frac{1}{2L+1} \sum_{|n| \leq L, n=2^m} q[2^m n] q[n - k] \]

\[ \quad = \lim_{L \to \infty} \frac{1}{2L+1} \sum_{|l| \leq L} q[l] q[2^m l - k]. \]

Since \( q[n] \) is correlation-ergodic, we may replace this correlation with its expected value:

\[ R_q[k] = \lim_{L \to \infty} \frac{1}{2L+1} \sum_{|l| \leq L} \delta[(2^m - 1)l - k] = \begin{cases} \delta[k] & m = 0 \\ 0 & \text{otherwise} \end{cases}. \]

Hence,

\[ S_{\omega,m}(\omega) = \begin{cases} 1 & m = 0 \\ 0 & \text{otherwise} \end{cases}. \]

where, without loss of generality, we have set \( \sigma^2 = 1 \). Then, using

\[ S_{\omega,m}(\omega) = (2\beta)^{-m/2} \Psi(2^{m}\omega) \Psi^{*}(2^{m}\omega) S_{\omega,m}(\omega) \]

we get that

\[ S_{\omega,m}(\omega) = \begin{cases} |\Psi(\omega)|^2 & m = 0 \\ 0 & \text{otherwise} \end{cases}. \quad (D.19) \]

Finally, we note that

\[ S_{\omega,m}(\omega) = S_{\omega,m}(\omega) \]

and that

\[ S_{\omega,m}(\omega) = j^{-m} S_{\omega,m}(2^{-m} \omega). \]

Using these identities together with (D.19) in (D.18) yields

\[ S(\omega) = \sum_{m} \beta^{-m} |\Psi(2^{-m}\omega)|^2 \]

as desired.

References


<table>
<thead>
<tr>
<th>References</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>References</th>
</tr>
</thead>
</table>

References


References


References


Index

Barnes-Allan model, 37, 42
Cramer-Rao bounds, 69
correlation function, 45
distribution of time-constants model, 43
elements, 33
fractal dimension, 33
fractal dimension estimation, 64
frequency-domain characterization, 41
Karhunen-Loève-type expansions, 46, 50
Keshner model, 44
maximum likelihood parameter estimation, 63
measured power spectra, 32
parameter estimation, 120
pole-zero behavior, 45
RC transmission line model, 44
sample paths, 32
smoothing (noise reduction), 72
spectrum estimation, 64
state variables, 45
Van der Ziel model, 43
wavelet coefficient correlation, 50-51
wavelet coefficient stationarity properties, 50
wavelet coefficient variances, 51
wavelet coefficients, ideal bandpass basis, 52
wavelet-based analysis, 50
wavelet-based models, 46
wavelet-based synthesis, 47
wavelet-based synthesis, regularity issues, 49
whitening filter, 61-62

A
autoregressive moving-average (ARMA) models, 35, 43

B
bandpass filter
ideal, continuous-time, 11
Barnes-Allan process, 37, 42, 45
bihomogeneous signal, 95

C
characteristic sequence, 108
synthesis, 110
communication channel models, 113
conjugate quadrature filter (CQF), 25-27
correlation function, 38
correlation structure, 39
differentiation in scale, 135
discrete wavelet transform, 21
analysis algorithm, 22
complexity, 22
data length and resolution effects, 23
synthesis algorithm, 22

D
differentiation in scale, 135
discrete wavelet transform, 21
analysis algorithm, 22
complexity, 22
data length and resolution effects, 23
synthesis algorithm, 22

E
Euler-Cauchy equations, 136

F
filter bank
octave band, 13
Fourier transform
definition, continuous-time, 10
definition, discrete-time, 21
fractal dimension, 33
estimation, 64
estimator performance, 71
estimator performance bounds, 69
fractal modulation
analog data, 124
choice of fractal dimension, 118
digital data, 121
discrete-time implementation, 117, 122-124
finite-length messages, 118
performance, 122, 126
power efficiency, 123
receiver, 120
requirements, 114
spectral efficiency, 116
spectral matching, 120
transmitter, 115
fractional Brownian motion, 30, 36
correlation function, 38
definition, 38
derivative, 39
increments, 38
limitations, 40
self-similarity parameter, 38
synthesis, 39, 139
uniqueness, 40
variance, 40
fractional calculus, 45
fractional Gaussian noise, 36
correlation structure, 39

G
generating sequence, 99

H
highpass filter
ideal, discrete-time, 22
homogeneous signal, 95
bihomogeneous signal, 95
characteristic sequence, 108
discrete-time, 109
energy-dominated, 96, 97
equivalence class, 98
examples, 97, 105
Fourier transform, 97, 103
fractal dimension, 104
generating sequence, 99
Hilbert space, 100
input to LTSI(X) system, 138
periodicity-dominated, 107, 118
power spectrum, 103-105
power-dominated, 96, 103
randomly-generated, 105
regularity, 104
time-frequency portrait, 107
wavelet expansion, 98

I
infrared catastrophe, 35
iterated function system, 110
J
Johnson noise, 60

L
Laplace transform, 130-132
linear scale-invariant system, 132
convolution property, 133
eigenfunctions, 133
generalized class, 136
lagged-impulse response, 132
system function, 134
linear system, 131
linear time- and scale-invariant system, 137
dyadic approximation, 143
wavelet representation, 140
linear time-invariant system, 131
impulse response, 131
long-term dependence, 39
Lorenzian spectra, 43

M
Mellin transform, 130, 134
inverse formula, 134
multiresolution signal analysis, 16
characterization, 17
detail signal, 20
perfectly bandlimited, 19

N
nearly-1/f process
spectrum, 47
wavelet-based synthesis, 47
nearly-1/f processes, 30, 43

P
Poisson formula, 12

R
regularity, 12

S
scale-differential equations, 135
scaling function, 18
ideal lowpass, 19
self-similar basis, 96, 99
example, 101
self-similar system, 130
shot noise, 60
spectral efficiency, 114
stable distributions, 35
statistical self-similarity, 31

T
thermal noise, 60
time-frequency analysis, 14

U
ultraviolet catastrophe, 35

W
wavelet
basic, 11
basic (or "mother"), 20
Battle-Lemarie, 27
Daubechies, 27
father, 18
frame, 10
Haar, 26
ideal bandpass, 11, 13, 15, 20,
26, 28
mother, 11
scaling function, 18
wavelet bases, 9
Battle-Lemarie, 27
construction of, 25
dyadic, 11
example lacking multiresolution
analysis, 12
examples, 26
filter bank interpretation, 13
multiresolution analysis inter-
pretation, 16
nondyadic, 27
orthonormal, 11

white Gaussian noise, 31, 35
Wiener process, 31, 35, 38