5-2013

Statistical Methods of Wavelet Analysis with Applications to Ecological Time-Series

Daniel J. Peach
Bates College, dpeach@bates.edu

Statistical Methods of Wavelet Analysis with Applications to Ecological Time-Series

Daniel Peach

Department of Mathematics

Bates College

Lewiston, ME 04240

Statistical Methods of Wavelet Analysis with Applications to Ecological Time-Series

An Honors Thesis

Presented to the Department of Mathematics

Bates College

in partial fulfillment of the requirements for the

Degree of Bachelor of Arts

by

Daniel Peach

Lewiston, Maine

March, 2013

Contents

Acknowledgments

Thank you for reading my thesis. It's the culmination of a year-long effort to get to know a little bit about a rapidly expanding new mathematical discipline - wavelets. Inside, you'll find both discrete and continuous wavelet theory, some programming, and, as it turned out, a lot of statistics.

This thesis is a little less focused than I would have hoped at the outset. It's a collection of ideas about how to approach a too-broad question: how can we use wavelets to tackle questions about ecological data sets? As I complete this thesis, I'm certain I didn't answer this question as completely as I could have. Nevertheless, I am glad to have had the opportunity to teach myself so much, and I think my advisor, Meredith Greer, will be able to use my findings to aid her research in the future.

I'd also like to thank Meredith for being a fantastic thesis advisor this year; without her coaching and careful editing, this thesis wouldn't have happened.

Daniel Peach

Introduction

0.1. Signals. The primary concern of this thesis will be the application of wavelet transforms to the analysis of ecological signals. It makes sense, then, to first establish what we mean by a *signal*. Consider the following non-mathematical definition:

sig·nal *n.* A gesture, action, or sound that is used to convey information or instructions.

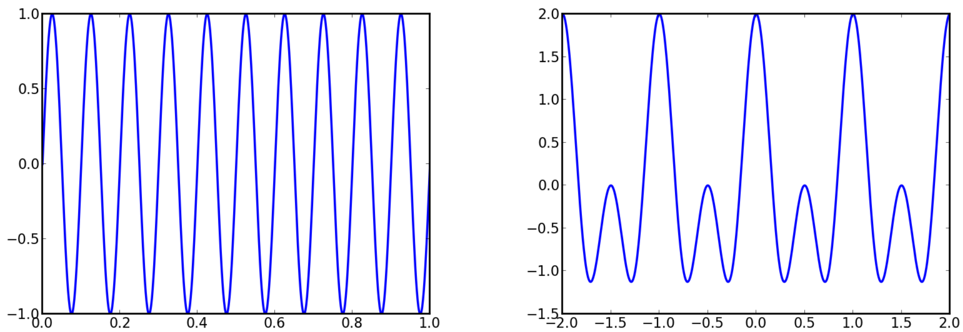
We characterize, model, and represent many phenomena—both naturally occuring and man-made—with signals: an electrocardiogram is a discrete, finite length, one-dimensional signal that represents a process occurring inside of a human body. Similarly, computer images are our most familiar examples of finite length, discrete, two-dimensional signals.

Mathematically, a signal is simply a function. Though we will place certain restrictions upon the functions we will allow as signals, the principal differentiating feature of signals—as indicated by the non-mathematical definition given above—is the emphasis we place upon the information content we associate with them. Signal processing, then, is the discipline that attempts to mathematically define and extract this information from signals.

In general, we ask two kinds of questions about signals:

(1) What kind of information can we extract from a given signal?
(2) Does some (hopefully small) set of data uniquely characterize a given signal – or differentiate it from other signals?

Answers to the questions posed above often use *frequency*[1] to classify signals. If we make certain assumptions about the kinds of signals we're working with, these questions given above turn out to be fairly simple to resolve.



(a) $f(x) = \sin(20\pi x)$ (b) $g(x) = \cos(2\pi x) + \cos(4\pi x)$

FIGURE 0.1

For example, suppose we want to tell someone about the signal shown in Fig. 0.1.(a). We could, of course, send them a picture of the signal. It's much more efficient, however, to simply tell them the information we can glean from simple observation: it's a 10 Hz signal with peaks at one and negative one.

Signals like $f(x)$ given in Fig. 0.1.(a). are called *stationary* signals; that is, these signals do not change in frequency or amplitude over time. *If* we're working with stationary signals, then the Fourier transform—the classical technique of signal processing—can give us information about a given signal that *uniquely* characterizes that signal. The Fourier transform—whether continuous or discrete—is a function of functions: it takes a function as its input and returns a new function as its output. The values of the output function correspond to the magnitudes of the component frequencies contained within the input function. In less mathematical terms, the information provided by the Fourier transform will resemble the sentence I used previously to describe Fig. 0.1.(a); that is, the Fourier transform tell us the frequencies and amplitudes contained within a stationary signal. Furthermore,

[1]Frequency is the rate at which a signal repeats itself. If a signal repeats itself once every second, then we say it is a 1 Hz signal; the signal given by Fig. 0.1.(a). is a 10 Hz signal. The units, of course, don't have to be in seconds - in fact, they don't even have to be units of time. Furthermore, a signal can have many constituent frequencies – see Fig. 0.1.(b).

the Fourier transform can tell us information about and let us mathematically manipulate more complicated signals—like Fig. 0.1.(b)[2].—or even noisy signals.

It turns out that we can safely assume that many signals are stationary: chemists and physicists, for example, will recognize that spectroscopy relies upon the stationary electromagnetic signals emitted by excited atoms. For these reasons—and a number of other reasons I've left unstated—the Fourier transform has been the ubiquitous analytical tool in science and applied mathematics for more than a hundred years.

Yet it's an imperfect assumption. It's not hard to see that stationary signals make up a fairly small subset of all mathematically possible signals: consider human speech, population growth, and, in the two-dimensional case, images[3]. Let's clarify the difference between stationary and non-stationary signals and demonstrate some of the shortcomings of Fourier analysis. Consider the following two signals, plotted in Figure 0.2:

(1) $f(x) = e^{-x^2}\left[\sin(2\pi x) + \sin(20\pi x) + \sin(10\pi x)\right]$

(2) $g(x) = e^{-(x+1)^2}\sin(2\pi x) + e^{-x^2}\sin(20\pi x) + e^{-(x-1)^2}\sin(10\pi x)$

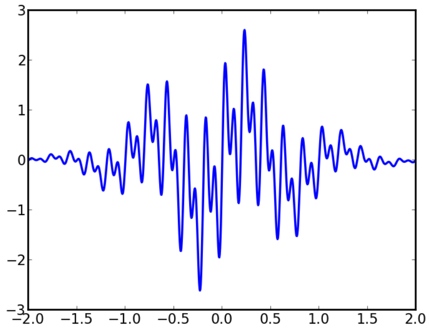
Qualitatively, the graphs given by $f(x)$ and $g(x)$ do not resemble one another. Yet the classical means of analyzing them—the Fourier transform[4]—cannot meaningfully differentiate between them either qualitatively or, as it turns out, quantitatively. At least when considering these two signals, the Fourier transform[5] fails to adequately resolve the two central questions of signal processing that I posed above. It's clear, then, why we might need a new technique for analyzing non-stationary signals. The main concern of the expository section of this thesis will be a discussion of a recently developed technique called *wavelet analysis* used to gather the information content of non-stationary signals.

[2]Note: we actually say that $g(x)$ is a stationary signal, even though it is visually quite complex.

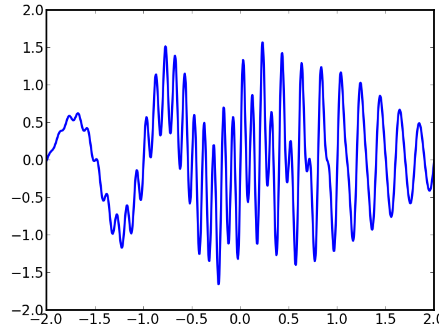
[3]Though it may not be obvious, an image is a finite two-dimensional signal; they can be analyzed with techniques that are generalizations of the techniques used for one-dimensional signals.

[4]A note on how to interpret the graphs given by (c). and (d): the horizontal axes indicate the component frequencies of $f(x)$ and $g(x)$, respectively.

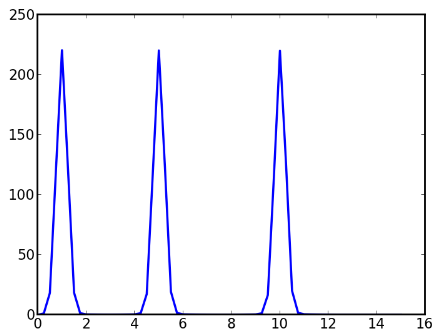
[5]For this thesis, I only refer to the discrete Fourier transform. There are different kinds of Fourier transforms, and while they each serve a different purpose, they all have the same shortcomings with respect to non-stationary signals.



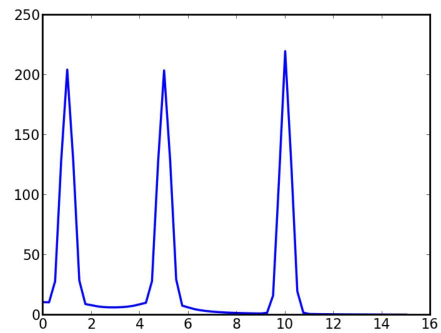
(a) $f(x)$



(b) $g(x)$



(c) Discrete Fourier transform of $f(x)$



(d) Discrete Fourier transform of $g(x)$

FIGURE 0.2. .

0.2. The Discrete Wavelet Transform. As we have just discussed, the wavelet transform and the Fourier transform are techniques used to analyze the frequency content of signals. The expository sections of this thesis will provide an introduction to the wavelet transform with a focus on the discrete wavelet transform (DWT). However, for reasons that I hope will become more clear shortly, it is important to define and articulate basics of the Discrete Fourier transform (DFT) so that it may be used as an expository, analytical, and computational tool.

There are plenty of conceptual reasons for introducing the DFT before the DWT; both transforms decompose a signal into its component frequencies, but the DFT does this in a much more intuitive and elementary way. For a reader unfamiliar with function transformations, the DFT will make the process of learning the DWT more palatable.

Furthermore, as we begin to examine the DWT, the frequency content of a given signal will become *defined* as the values registered by the DFT; we will use the DFT as an analytical tool to construct and evaluate the wavelets used in the DWT. Finally, we will prove that we can compute the DWT by computing of a series of convolutions. Those familiar with the Fourier transform will know of the intimate relationship between the Fourier transform and convolution; the use of the Fast Fourier transform (FFT) will allow us to quickly compute the DWT.

Chapter 1 contains the background linear algebra necessary to understand the DFT; it also contains the statistics necessary for later chapters. Within Chapter 2, I will introduce the DFT and prove some important results that characterize the transform; in particular, I will introduce the convolution operator and frame its relationship to the DFT. Finally, in Chapter 3, I will begin my introduction to the DWT; as highlight, I introduce two new statistical significance tests for the Haar wavelet transform. Chapter 5 will include applications of these significance tests to ecological signals.

As a final note, most of the definitions and proofs in these first chapters are adapted from Michael Frazier's text *An Introduction to Wavelets Through Linear Algebra*, although I have expanded or corrected them when they are too terse or unclear. I will note when the material does not come from Frazier's text.

0.3. The Continuous Wavelet Transform. I want to explain why this thesis is more disjointed than I would have hoped, and I also want provide a picture of what the field of wavelet analysis really looks like. I would have greatly appreciated this kind of an outline of wavelet theory before I began this thesis.

When I began this thesis, I knew that I would be applying the techniques I learned to discrete signals; I assumed, therefore, that I should focus on the DFT and the DWT. Surprisingly enough, I was looking at the wrong transforms - at least for the kind of signal analysis that I would end up really focusing on. The DWT is a fantastic tool for certain kinds of fast, technical processing - it's good for image compression, pattern recognition, and signal denoising - but it's not good for what I wanted to focus on.

The *continuous* and *discrete* wavelet transforms do not constitute the entirety of wavelet analysis; in fact, there exists an entire field of wavelet analysis that relies upon a structure called a *multiresolution analysis.* This field, though it uses continuous wavelet functions, is actually most closely related to the DWT in its organization; in Fourier analysis, its closest analogue is the Fourier series. It is also the most sophisticated and theoretically developed field within wavelet analysis; its most prominent applications are in physics and in differential equations.

Both the DWT and the multiresolution analysis-reliant wavelet transforms are concerned with the orthogonal decomposition of a signal. While orthogonal decompositions have many virtues, they are not always the best scientific tool. In ecological applications, our primary concern is exactly locating events in time and frequency within a signal; it turns out that orthogonal decompositions actually obfuscate this information by diffusing it across a fairly small number of basis vectors[6]. We will use a discretely sampled continuous wavelet transform (CWT) to analyze ecological signals; this will give us the most information possible about a given signal. Thus, even though the CWT is the least sophisticated - and most computationally intensive - of the wavelet transforms, it is the appropriate tool for the job.

The primary mathematical challenge of this thesis will be the construction and implementation of statistical tests. When we take the CWT of white noise[7], we see apparently cohesive structures - that is, structures that resemble the wavelet transforms of sinusoids or other traditional functions. Thus, when we don't *a priori* know the content of a signal, we need statistical tests to identify white noise.

Chapter 4 will include an introduction to the CWT; we will also derive and alter established statistical significance tests that use the CWT. Chapter 5 will include applications of these significance tests to ecological signals.

[6]It's hard to explain this in a non-technical way. For those who are familiar with this material, consider what happens when we take the DFT of a signal whose component frequencies do not exactly match the frequencies of the DFT basis vectors – the frequency will be "shared" by nearby vectors, and we won't be able to exactly determine the component frequencies of our original signal. (Actually, the problem is much worse than that – it's diffused across *all* of the basis vectors. We will discuss this problem in detail.)
[7]That is, a signal produced by randomness. We will define this more carefully later on.

CHAPTER 1

Background Linear Algebra and Statistical Concepts

1. Linear Algebra

The first three chapters of this thesis will introduce two discrete transforms. This section aims to introduce the linear algebra requisite for understanding both the DFT and, later on, the DWT.

In rough lay language, we say that the Fourier transform decomposes a function into its component frequencies. In the more precise terminology of linear algebra, the Fourier transform projects a function onto a set of orthogonal basis vectors whose properties we understand in terms of frequency; the magnitude of the weights on these basis vectors corresponds to the presence of a certain frequency in a function. The goal of this chapter is to introduce the mathematical background necessary to understand the concept of a projection of a vector onto an orthogonal vector space; the Fourier transform will follow straightforwardly from these concepts in Chapter 2.

1.1. Basic Structures and Definitions. This thesis will assume familiarity with the properties of complex numbers and some basic facts of linear algebra. We begin by introducing the definition of a *vector space*; this will be the underlying structure of the Fourier and wavelet transforms.

Definition 1. Let $\mathbb{F}$ be a field. A **vector space** V over $\mathbb{F}$ is a set with operations of vector addition $+$ and scalar multiplication $\cdot$ satisfying the following properties:

(1) For all $u, v \in V$, $u + v$ is defined and is an element of V. (Closure for addition)
(2) For all $u, v \in V$, $u + v = v + u$. (Commutativity for addition)
(3) For all $u, v, w \in V$, $u + (v + w) = (u + v) + w$. (Associativity for addition)

(4) There exists an element in V, denoted as 0, such that $u + 0 = u$ for all $u \in V$. (Additive identity)

(5) For each $u \in V$ there exists an element in V, denoted $-u$, such that $u + (-u) = 0$ (Additive inverse).

(6) For all $\alpha \in \mathbb{F}$ and $u \in V$, $\alpha \cdot u$ is defined and is an element of V. (Closure for scalar multiplication)

(7) For all $u \in V$, $1 \cdot u = u$, where 1 is the multiplicative identity in $\mathbb{F}$. (Scalar multiplicative identity)

(8) For all $\alpha, \beta \in \mathbb{F}$ and $u \in V$, $\alpha \cdot (\beta \cdot u) = (\alpha\beta) \cdot u$. (Associativity for scalar multiplication)

(9) For all $\alpha \in \mathbb{F}$ and $u, v \in V$, $\alpha \cdot (u + v) = (\alpha \cdot u) + (\alpha \cdot v)$. (First distributive property)

(10) For all $\alpha, \beta \in \mathbb{F}$ and $u \in V$, $(\alpha + \beta) \cdot u = (\alpha \cdot u) + (\beta \cdot u)$. (Second distributive property)

The set of n-dimensional Euclidean vectors $\mathbb{R}^n$ is our most familiar example of a vector space. However, we can apply the definition of a vector space to diverse sets of mathematical objects.

Example 1. All of the following sets are vector spaces.

(1) The set of $n-1$ degree polynomials, $\mathbb{P}_n$, under polynomial addition is a vector space.

(2) The set of $m \times n$ matrices, $\mathbb{R}^{m\times n}$, under matrix addition is also vector space.

(3) The set of square-integrable functions under point-wise addition:

$$L^2(\mathbb{R}) = \left\{ f : \mathbb{R} \to \mathbb{C} : \int_{-\infty}^{\infty} |f(x)|^2 < \infty \right\}.$$

(4) The set of square-summable sequences under point-wise addition:

$$\ell^2(\mathbb{R}) = \left\{ a_n : \mathbb{R} \to \mathbb{C} : \sum_{n=0}^{\infty} |a_n|^2 < \infty \right\}.$$

For our purposes, the underlying field we will use for our vector spaces will be $\mathbb{C}$, the set of complex numbers.

Definition 2. We define the vector space $\ell^2(\mathbb{Z}_N)$ as

$$\ell^2(\mathbb{Z}_N) = \{z = (z(0), z(1), ..., z(N-1)) : z(j) \in \mathbb{C}, 0 \leqslant j \leqslant N-1\}.$$

That is, $\ell^2(\mathbb{Z}_N)$ is the set of complex valued functions from $\mathbb{Z}_N$.

The vector space $\ell^2(\mathbb{Z}_N)$ will be the focus of this thesis. Although we have defined each vector in $\ell^2(\mathbb{Z}_N)$ as the ordered values of a function from $\mathbb{Z}_N$ to $\mathbb{C}$, we should note that, in general, an element of $\ell^2(\mathbb{Z}_N)$ will look and mathematically act like an element of $\mathbb{C}^N$. Yet we should remember to interpret an element $z \in \ell^2(\mathbb{Z}_N)$ as a signal: in most applications, the values of z will be discrete samples taken at some constant sampling rate.

While we often think of a function in $\ell^2(\mathbb{Z}_N)$ as a finite length, we note that the domain of $\ell^2(\mathbb{Z}_N)$ is $\mathbb{Z}_N$, the set of integers modulo N. Later on, we will exploit this fact: a function in $\ell^2(\mathbb{Z}_N)$ is actually N-periodic.

Consider arbitrary vectors $u, v, w \in \mathbb{R}^2$. If we can find some $\alpha, \beta \in \mathbb{R}$ such that $u = \alpha v + \beta w$, then we say that u is a **linear combination** of v and w. We can also consider the set of all linear combinations of v and w.

Definition 3. Let V be a vector space over a field $\mathbb{F}$, and suppose $U \subseteq V$. The span of U (denoted span U) is the set of all linear combinations of U. In particular, if U is a finite set, say

$$U = \{u_1, u_2, ..., u_n\},$$

then

$$\text{span } U = \left\{ \sum_{j=1}^{n} \alpha_j u_j : \alpha_j \in \mathbb{F} \text{ for all } j = 1, 2, ..., n \right\}.$$

Example 2.

Let X be the set of all vectors in $\mathbb{R}^2$ lying on the line $y = x$. Then

$$X = \text{span } \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}.$$

That is, when we consider all of the linear combinations of the vector $(1,1)$ in $\mathbb{R}^2$, we get *every* vector lying on the line $y = x$.

Example 3.

$$\mathbb{R}^2 = \text{span}\left\{\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix}\right\} = \text{span}\left\{\begin{pmatrix}1\\2\end{pmatrix},\begin{pmatrix}0\\\sqrt{3}\end{pmatrix}\right\}$$

Both sets of vectors given above span $\mathbb{R}^2$; that is, we can rewrite *any* vector in $\mathbb{R}^2$ as a linear combination of the vectors in those sets. We can generalize this concept for arbitrary vector spaces.

Definition 4. Let V be a vector space over a field $\mathbb{F}$. A subset U of V is a **basis** for V if U is a linearly independent set such that span $U = V$.

Example 4. For $\mathbb{R}^n$, the **standard basis** is given by $\{\mathbf{e}_i\}_{i=1}^n$, where $\mathbf{e}_i \in \mathbb{R}^n$ and $e_i = 1$ for its i^{th} coordinate and 0 elsewhere.

As seen in Example 3, there can be different bases for a single given vector space; it turns out that for an n dimensional vector space V, any set of n linearly independent vectors $\{v_i\}_{i=1}^n \in V$ will be a basis of V.

It may not be obvious why we want to find bases for a vector space; it may be even less obvious why we might choose one basis over another. We are not yet ready to address these concerns, but for now, note that, in general, we tend to choose bases in accordance with the mathematical problems we are trying to solve. Although the reader has no reason to understand what this might mean at this point, it should be noted that the difference between Fourier analysis and wavelet analysis really comes down to a choice of basis.

For a given vector $v \in V$, we can "locate" v with respect to our basis vectors.

Definition 5. Suppose V is a vector space over a field $\mathbb{F}$ and $S = \{v_1, v_2, ..., v_n\}$ is a basis for V. For any vector $v \in V$, there exist unique $\alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{F}$ such that $v = \sum_{j=1}^n \alpha_j v_j$. We say that $[v]_S$ is the **coordinate vector of v with respect to S**; we write $[v]_S$ as the vector in $\mathbb{F}^n$ with components $\alpha_1, \alpha_2, ..., \alpha_n$, that is,

$$[v]_S = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

EXAMPLE 5.

Let $B = \left\{ \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$. It is straightforward to show that B is a basis for $\mathbb{R}^2$.

Then let $v = \begin{pmatrix} 7 \\ 3 \end{pmatrix}$. Then $[v]_B = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

1.2. Projection. For an arbitrary element $v \in V$ and a given basis B, how can we determine $[v]_B$? That is, how can we determine $\{\alpha_i\}_{i=1}^n$? It turns out that we generally avoid answering this question for arbitrary bases; we restrict our analyses to *orthogonal* bases. To understand what this means—and why we make this restriction—we first need to introduce a few more definitions.

DEFINITION 6. Let V be a vector space over $\mathbb{C}$. A **complex inner product** is a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ with the following properties:

(1) For all $u, v, w \in V$, $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$.
(2) For all $\alpha \in \mathbb{C}$ and all $u, v \in V$, $\langle \alpha u, v \rangle = \alpha \langle u, v \rangle$.
(3) For all $u, v \in V$, $\langle u, v \rangle = \overline{\langle v, u \rangle}$.
(4) For all $u \in V$, $\langle u, u \rangle \geqslant 0$, and $\langle u, u \rangle = 0$ if and only if $u = 0$.

We call a vector space associated with a complex inner product a **complex inner product space**.

EXAMPLE 6. Let $x, y \in \mathbb{C}^n$. Let x_i denote the i^{th} element of x. Then the **dot product**

$$x \cdot y = \sum_{i=1}^{n} x_i \overline{y_i}$$

is an inner product.

Recall the following identity:

$$x \cdot y = ||x||||y|| \cos\theta,$$

where θ is the angle between x and y.

Consider $x_1, x_2, y \in \mathbb{R}^2$ where $||x_1|| = ||x_2|| = 1$. Suppose we want to write y as a linear combination of x_1 and x_2; that is,

$$y = \alpha x_1 + \beta x_2$$

for some $\alpha, \beta \in \mathbb{R}^2$. Since

$$x_1 \cdot y = ||y|| \cos\theta,$$

it's straightforward to see that, geometrically (and trigonometrically[1]), the dot product measures "how much" of y points in the x_1 direction.

We might think that we can just let $\alpha = x_1 \cdot y$ and $\beta = x_2 \cdot y$. Unfortunately, it usually doesn't work out that easily: in most cases, there's a bit of "overlap" between x_1 and x_2—that is, $x_1 \cdot x_2 \neq 0$—and this makes the computation of α and β much more difficult[2]. It turns out that we can let $\alpha = x_1 \cdot y$ and $\beta = x_2 \cdot y$ only in the case that $x_1 \cdot x_2 = 0$.

We can generalize (and prove) this whole process for an arbitrary inner product space.

Definition 7. Let V be a complex inner product space. Then we define the **norm** of $u \in V$ as

$$||u|| = \sqrt{\langle u, u\rangle}.$$

We now introduce the concept of *orthogonality*; orthogonality generalizes the notion of "overlap" (more exactly, a lack of "overlap") between vectors.

[1]Try drawing a picture as an example.

[2]Again, try drawing a picture. Here's the problem: if $x_1 \cdot x_2 \neq 0$, then x_1 points at least a little bit in the x_2 direction (or vice versa, depending on your perspective). Remember that $x_1 \cdot y$ tells us how much y points in the x_1 direction and $x_2 \cdot y$ tells us how much y points in the x_2 direction. But since $x_1 \cdot x_2 \neq 0$, $x_1 \cdot y$ also tells us a *little* bit about how much x_2 points in the y direction, which was supposed to be already accounted for by $x_2 \cdot y$.

Definition 8. Suppose V is a complex inner product space. For $u, v \in V$, we say that u and v are **orthogonal** if $\langle u, v\rangle = 0$.

Example 7.

Let $x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, and $y = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Then since

$$x \cdot y = \sum_{i=1}^{2} x_i \overline{y_i} = 1 \cdot 0 + 0 \cdot 1 = 0,$$

x and y are orthogonal vectors.

Definition 9. Suppose V is a complex inner product space. Let B be a collection of vectors in V. We say that B is an **orthogonal set** if any two different elements of B are orthogonal. Furthermore, B is an **orthonormal set** if B is an orthogonal set and $||v|| = 1$ for all $v \in B$.

Example 8. For $\mathbb{R}^n$, the standard basis is an orthonormal basis.

The following theorem exactly generalizes our previous discussion of vectors in $\mathbb{R}^2$.

Theorem 1.1. Suppose V is a complex inner product space. Let $B = \{v_1, v_2, ..., v_n\}$ be an orthogonal basis for V. Let $u \in V$. Then

$$u = \sum_{i=1}^{n} \frac{\langle u, v_i\rangle}{\langle v_i, v_i\rangle} v_i. \tag{1.1}$$

Furthermore, if B is an orthonormal basis for V, then

$$u = \sum_{i=1}^{n} \langle u, v_i\rangle\, v_i. \tag{1.2}$$

Proof. If $u \in V$, then $u \in \text{span } B$. Therefore, there exists a set of scalars $\{\alpha_i\}_{i=1}^{n}$ such that

$$u = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n.$$

For each i, to determine α_i, we take the inner product of both sides of the equation and v_i. That is,

$$\begin{aligned}\langle u, v_i\rangle &= \langle \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n, v_i\rangle \\ &= \langle \alpha_1 v_1, v_i\rangle + \langle \alpha_2 v_2, v_i\rangle + \ldots + \langle \alpha_n v_n, v_i\rangle \\ &= \alpha_1 \langle v_1, v_i\rangle + \alpha_2 \langle v_2, v_i\rangle + \ldots + \alpha_n \langle v_n, v_i\rangle .\end{aligned}$$

But since B is an orthogonal basis, $\langle v_j, v_k\rangle = 0$ for $j \neq k$. Therefore,

$$\langle u, v_i\rangle = \alpha_i \langle v_i, v_i\rangle .$$

Solving for each α_i yields the desired result.

Furthermore, if B is an orthonormal basis, then $\langle v_i, v_i\rangle = 1$; therefore, $\alpha_i = \langle u, v_i\rangle$. □

At this point, it's not hard to see that if B is an orthonormal basis, then

$$[u]_B = \begin{pmatrix} \langle u, v_1\rangle \\ \langle u, v_2\rangle \\ \vdots \\ \langle u, v_n\rangle \end{pmatrix}.$$

This is called the **projection** of a vector onto an orthogonal vector space; we're going to be using that terminology a lot, mostly because the word is so evocative of what's actually going on.

It's important to consider the use of the definition of orthogonality in our proof. Mathematically, the fact that we used an orthogonal basis greatly simplified the computation of $[u]_B$. Conceptually, however, we should consider that orthogonality implies that there is no "overlap" between our basis vectors: if we project a vector u onto an orthogonal basis, then we are finding exactly "how much" of each basis vector is contained within u.

We will finish this section with a theorem from linear algebra that will be useful for us later; conceptually, it is relatively unimportant.

THEOREM 1.2. Suppose V is a complex inner product space. Suppose B is an orthogonal set of vectors in V and $0 \notin B$. Then B is a linearly independent set.

PROOF. Suppose $u_1, u_2, ..., u_k \in B$ and there exist scalars $\alpha_1, \alpha_2, ..., \alpha_k$ such that

$$\alpha_1 u_1 + \alpha_2 u_2 + \ldots + \alpha_k u_k = 0.$$

Take the inner product of both sides with u_j for $j \in \{1, 2, ..., k\}$. Since B is an orthogonal set, $\langle u_i, u_j \rangle = 0$ for $i \neq j$. Then

$$\alpha_j \langle u_j, u_j \rangle = 0.$$

But since $u_j \neq 0$ by assumption, $\langle u_j, u_j \rangle \neq 0$. Therefore, $\alpha_j = 0$. Since j is arbitrary, this proves that B is a linearly independent set. □

2. Statistical Concepts

We will need some fairly sophisticated statistical tools for the later sections of this thesis; this section will act as a review for those who have a basic knowledge of probability. In particular, we will assume an understanding of **random variables** and **independence** of random variables. These definitions and theorems in this section are taken from [**5**]. Though this chapter should be used as a reference for the rest of the thesis, some concepts and theorems are introduced *ad hoc* throughout the thesis for clarity.

In general, we will use the common " $\sim$ " convention for denoting the probability density of a random variable; for example, if X is a random variable from a standard normal distribution, we will write

$$X \sim N(0, 1).$$

Furthermore, when we talk about random variables, we'll always use capital letters: e.g., we'll say that X or Y is a random variable. However, if we're talking about an observed value of a random variable, we'll use lower-case letters: e.g., the random variable X took on the value $x = 2$.

As a final note, we haven't proved a lot of the theorems in the upcoming sections, especially theorems about functions of random variables. Proofs of these theorems require moment-generating functions, but they also require facts from analysis about the relationship between moment-generating functions and their associated probability density functions. We had to cut things off at some point; here's where we drew the line.

2.1. Important Probability Densities.

DEFINITION 10. A random variable X has a **normal distribution** if its probability density is given by

$$N(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \text{ for } -\infty < x < \infty,$$

where $\sigma > 0$.

We call a random variable with probability density function $N(x;0,1)$ a **standard normal distribution**.

EXAMPLE 9. Let X be a random variable with density function $N(x;0,1)$. What's the probability that X takes a value between -1 and 1?

Those familiar with continuous random variables should know that we answer this kind of question by integrating the probability density function between the values of interest. That is,

$$P(-1 < X < 1) = \int_{-1}^{1} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(x)^2} \approx 0.682.$$

Note that there isn't an analytic solution to the integral of the normal distribution - we calculate the integral numerically.

DEFINITION 11. We denote a white noise signal by $W \in \ell^2(\mathbb{Z}_N)$. For $i = 0, 1, ..., N-1$, $W(i)$ has the following properties:

(1) The random variable $W(i)$ is normally distributed; that is, $W(i) \sim N(\mu,\sigma)$.

(2) For all $i \neq j$, $W(i)$ and $W(j)$ are independent.

DEFINITION 12. A random variable X has a **chi-square distribution** if its probability density is given by

$$f(x) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\frac{\nu-2}{2}} e^{-\frac{x}{2}} & \text{for } x > 0 \\ 0 & \text{elsewhere,} \end{cases}$$

where Γ is the **Gamma function** and $\nu \in \mathbb{N}$ is the **degrees of freedom** of X. We denote a random variable with a χ^2 distribution with ν degrees of freedom by $X \sim \chi^2_\nu$.

Those unfamiliar with the Gamma function need not worry - like the normal distribution above, we will only work with the chi-square distribution numerically.

2.2. Moments of Random Variables. We use *moments* of random variables to quantify the kinds of values we expect to see from them.

DEFINITION 13. If X is a continuous random variable and $f(x)$ is the value of its probability density at x, the **expected value** of X is

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x f(x) dx.$$

THEOREM 1.3. Let g be a continuous function. If X is a continuous random variable and $f(x)$ is the value of its probability density at x, the expected value of $g(X)$ is given by

$$\mathbf{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx.$$

THEOREM 1.4. For $a, b \in \mathbb{C}$,

$$\mathbf{E}[aX + b] = a\mathbf{E}[X] + b.$$

PROOF. The theorem demonstrates that expected value is a linear operator; the proof of the theorem follows directly from the definition of expected value, Theorem 1.3, and the fact that integration is a linear operator. $\square$

DEFINITION 14. The rth **moment about the origin** of a random variable X, denoted by μ'_r, is the expected value of X^r; that is, for $r = 0, 1, 2, ...$,

$$\mu'_r = \mathbf{E}[X^r] = \int_{-\infty}^{\infty} x^r f(x) dx.$$

where $f(x)$ is the probability density function of X. We call μ'_1 the **mean** of X; we often denote the mean with μ.

DEFINITION 15. The rth **moment about the mean** of a random variable X, denoted by μ_r, is the expected value of $(X - \mu)^r$; that is, for $r = 0, 1, 2, ...$,

$$\mu_r = \mathbf{E}[(X - \mu)^r] = \int_{-\infty}^{\infty} (x - \mu)^r f(x) dx.$$

where $f(x)$ is the probability density function of X. We call μ_2 the **variance** of X; we often denote the variance with σ^2 or by $\mathbf{Var}[X]$. Furthermore, we call the positive square root of the variance the **standard deviation** of X; we denote the standard deviation with σ.

THEOREM 1.5. Let X have a normal distribution with probability density $N(x; \mu, \sigma)$. Then X has mean μ and variance σ^2.

THEOREM 1.6. Let X be a random variable with a χ^2_ν distribution. Then X has mean $\mu = \nu$ and variance $\sigma^2 = 2\nu$.

For most applications, we're concerned with only three different moments of a given random variable: the mean, μ, the variance σ^2, and the second moment about the origin, μ'_2. Our interest in the second moment stems from the following theorem.

THEOREM 1.7.

$$\sigma^2 = \mu'_2 - \mu^2$$

PROOF.

$$\begin{aligned}\sigma^2 &= \mathbf{E}[(X-\mu)^2] \\ &= \mathbf{E}[X^2 - 2X\mu + \mu^2] \\ &= \mathbf{E}[X^2] - 2\mu\mathbf{E}[X] + \mu^2 \\ &= \mathbf{E}[X^2] - 2\mu^2 + \mu^2 \\ &= \mu_2' - \mu^2.\end{aligned}$$

□

Often, it's easier for us to compute μ_2' and μ^2 than to compute σ^2 directly; furthermore, when $\mu = 0$, $\sigma^2 = \mu_2'$.

2.3. Functions of Random Variables. The following theorems will be invaluable for us later on.

THEOREM 1.8. If X has a standard normal distribution, then X^2 has a chi-square distribution with $\nu = 1$ degrees of freedom.

THEOREM 1.9. If $X_1, X_2, ..., X_n$ are independent random variables having standard normal distributions, then

$$Y = \sum_{i=1}^{n} X_i^2$$

has a chi-square distribution with $\nu = n$ degrees of freedom.

THEOREM 1.10. Let $X_1, X_2, ...X_n$ be independent random variables; let $Y = \sum_{i=1}^{n} a_i X_i$. Then

$$\mathbf{Var}[Y] = \sum_{i=1}^{n} a_i \mathbf{Var}[X_i].$$

DEFINITION 16. If $X_1, X_2, ..., X_n$ are independent and identically distributed random variables, we say that they constitute a **random sample** from the infinite population given by their common distribution.

This is the standard way of describing a random sample, but the "identically distributed" part isn't straightforward to interpret. We say that the X_i's are **identically distributed** if their corresponding probability density functions $f_i(x)$ are equal for all i. We'll talk about this more in the next section, but this is a pretty big assumption to make.

DEFINITION 17. Let $X_1, X_2, ..., X_n$ be a random sample. Then we define the **sample mean** as

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and the **sample variance** as

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}.$$

As sums of random variables, both $\overline{X}$ and S^2 themselves are random variables; their distribution depends upon the distribution of the X_i's[3]. However, we can calculate important statistical moments associated with $\overline{X}$ and S^2 without knowing their probability density functions; these moments demonstrate their utility.

THEOREM 1.11. Let $X_1, X_2, ..., X_n$ be a random sample. Then

$$\mathbf{E}[\overline{X}] = \mu,$$

and

$$\mathbf{E}[S^2] = \sigma^2.$$

When we substitute x_i for X_i, $\overline{x}$ is what we usually think of as the "average" of a data set, where each x_i is an observation - e.g., x_i is the grade that the ith student received on an exam.

We can **standardize** any normally distributed random variable. If X is normally distributed with mean μ and variance σ^2, then

$$Z = \frac{X - \mu}{\sigma}$$

[3]When n is large, $\overline{X}$ is well approximated by a normal distribution; see the Central Limit Theorem.

has a standard normal distribution. In many cases, we don't know μ and σ^2. However, if $X_1, X_2, ..., X_n$ constitute a random sample of normally distributed random variables, then we normalize X_i such that

$$Z_i = \frac{X_i - \overline{X}}{S}$$

is approximately normally distributed for all i.

Procedure 1. **Normalization**. Let $x \in \ell^2(\mathbb{Z}_N)$. Then we normalize x with the following procedure:

(1) Compute $\overline{x}$, the sample mean of x.
(2) Find s^2, the sample variance of x.
(3) For all i, define $z \in \ell^2(\mathbb{Z}_N)$ as

$$z(i) = \frac{x(i) - \overline{x}}{s}$$

for $i = 0, 1, ..., N - 1$.

We will need this procedure for our tests later on.

2.4. Hypothesis Testing. In statistics, we often want to make claims about data sets produced by random variables (or, importantly, to claim that a data set *isn't* produced by a random variable; that is, the data set is produced deterministically). Unfortunately, it's hard to get a grasp on how to appropriately frame these kinds of decisions. I want to provide an example to clarify and motivate the need for the kind of decision-making structures we're going to introduce shortly.

Suppose you work for a company that produces 12 oz. cans of soda[4]. Your boss has asked you to carefully construct a statistical test to ensure that the company's factory is working perfectly. Before we can discuss hypothesis testing, we have to define what "working perfectly" might even mean.

[4]That is, aluminum cans with soda in them. Ignore the weight of the can. Just use 12 oz.

If a given soda can weighs more than 12 oz, then you're wasting soda; if it weighs less, then you're ripping off your customers. Your boss knows there are certain inevitable variations in soda can weight - machines and people aren't perfect, and a given soda can probably *won't* weigh exactly 12 oz. - but we won't say that the existence of these errors means that there's something wrong with the factory. In fact, we might even find that the average weight of all the soda cans ever produced by the factory isn't exactly 12 oz - *even if the factory is "working perfectly"*. What does "working perfectly" mean, then?

Let's say that the weight of the ith soda produced by your factory is governed by a random variable X_i. We assume that the X_i's are identically distributed, although this actually assumes a lot about your factory[5]. With that assumption out of the way, we'll start referring to a random variable X, which has the same distribution as all the X_i's - our X will be a representative soda can weight.

What your boss really means by the factory "working perfectly" is that X has mean $\mu = 12$ oz[6]. In essence, *hypothesis testing* is a framework for making decisions about whether your factory is working perfectly or not; that is, when can we say with confidence that X really has mean $\mu = 12$?

Let's say you carefully take a random sample of n soda cans from your factory. We know that if $\mu = 12$, then $\mathbf{E}[\overline{X}] = 12$. But what should we conclude if the observed mean isn't equal to twelve? For example, what should we conclude if $\overline{x} = 12.2$? Should our conclusion be different if $\overline{x} = 12.5$? At what point are the observed values so weird that we should decide that there's something wrong with the factory? That is, at what point should we decide that $\mu \neq 12$? Given our assumptions about the situation, computing the answer to these questions turns out to be trivial; conceptually, however, these are sophisticated

[5]Suppose there are two different sections of the factory that both produce soda cans. If one of the sections is more sloppy about its production process than the other, then your assumption that the X_i's are identically distributed populations is false. Nevertheless, if, in this example, we're working with the same kinds of machines over a fairly short period of time, then the assumption isn't a bad one.

[6]Your boss, if she's a smart boss, should also care about the variance σ^2 of X. If the variance is large, as the word evokes, there will be a lot of variation in soda can weight, even if the mean μ of X is equal to 12 exactly.

and hard-to-answer questions. To appropriately frame the question, we need to define a statistical hypothesis.

Definition 18. A **statistical hypothesis** is an assertion about the distribution of one or more random variables. If a statistical hypothesis completely specifies the distribution, it is called a **simple hypothesis**; otherwise, it is called a **composite hypothesis**.

Example 10. The following are all examples of statistical hypotheses.

(1) The random variable X has a normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$. (simple hypothesis)
(2) The random variable Y has a normal distribution. (composite hypothesis)
(3) The random variable Y does *not* have a normal distribution. (composite hypothesis)

We often compare two differing hypotheses. We often call a hypothesis that is widely accepted, or the current view, the **null hypothesis** or H_0; a more radical alternative claim is called the **alternative hypothesis** or H_A. If these definitions sound shaky to you, you're right. But in an applied context, it's usually clear what the null and alternative hypotheses actually are[7]. In the soda can factory example provided above, our null hypothesis is that X is normally distributed with mean $\mu = 12$; our alternative hypothesis is that X is normally distributed with mean $\mu \neq 12$.

Let's make a decision about our soda can factory based upon a single sample - we're just going to randomly pick and weigh a single soda can. Let's say that you decided before you picked your sample that you were going to reject your null hypothesis that $\mu = 12$ if and only if your sample weighs less than 11 oz. or more than 13 oz.

Definition 19. We identify the two error types associated with statistical decision making.

[7]We shouldn't think of the alternative hypothesis as the hypothesis that advances research: we try to stay objective with our language; also, in the example given above about the factory, we may hope that the null hypothesis is true - we don't want there to be something wrong with the factory.

(1) If we reject the null hypothesis, H_0, in favor of the alternative hypothesis H_A when the null hypothesis is, in fact, true, then we say that we have made a **Type I error**. We say that α is the probability of committing a Type I error.

(2) If we fail to reject the null hypothesis, H_0 when, in fact, the alternative hypothesis, H_A, is true, then we have made a **Type II error**. We say that β is the probability of committing a Type II error.

Example 11. In the factory example provided above, what's the probability of committing a Type I error? Note, of course, that to compute α we have to first identify the **rejection region** - the values of X that will cause us to reject the null hypothesis. In our example, we reject the null hypothesis when $X > 13$ or $X < 11$.

$$\alpha = 1 - P(11 < X < 13|\mu = 12) = 1 - \int_{11}^{13} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-12)^2} \approx .3173$$

This means that when we make decisions based upon the identified rejection region, we're going to mistakenly reject the null hypothesis about 32% of the time - *even though it is true.*

In the example above, we first identified the rejection region; we then identified the probability of committing a Type I error. We can also identify a good α first, and then compute an appropriate rejection region[8].

Example 12. Suppose X is a random variable. We want to test the null hypothesis that X is chi-square distributed with 2 degrees of freedom against the alternative hypothesis that X isn't chi-square distributed with 2 degrees of freedom[9]. If we let $\alpha = .05$, then what should our rejection region be?

We're looking for values of X that are unlikely to be seen from a chi-square distribution - we're looking to identify a rejection region. That is, we're looking for a k such that:

$$P(X \geqslant k|X \sim \chi^2_2) = \alpha.$$

[8]Note: there are infinitely many rejection regions corresponding to a single α. However, there are common-sense ways of choosing the appropriate rejection region based upon the claims you're trying to prove.
[9]Note: it's really hard to calculate β in this case. This is a huge problem with this kind of test.

This is equivalent to solving the following equation for k:

$$1 - \int_0^k \frac{1}{2\Gamma(1)} e^{-\frac{x}{2}} dx = \alpha.$$

With a little help from a numerical integrator and a equation solver, we find that $k \approx 5.99$.

Again, we should be careful to interpret this rejection region appropriately: it means that 95% of the time, a chi-square random variable with two degrees of freedom will take on values less than 5.99. Therefore, if we observe a random variable X take on a value greater than or equal to 5.99, then we can claim, with 95% confidence, that X isn't chi-square distributed with two degrees of freedom.

CHAPTER 2

Fourier Analysis

Chapter 1 introduced the projection of a vector onto an orthogonal basis. The goal of this chapter is to demonstrate that the Fourier transform is a projection of a certain kind of vector onto a certain kind of basis—and that we can interpret this basis in terms of frequency.

1. The Discrete Fourier Transform

The Discrete Fourier transform (DFT) projects discrete, finite length functions onto a discrete, finite length trigonometric basis. With this in mind, we now introduce the vector space with which we will be working for the remainder of this thesis.

THEOREM 2.1. For $u, v \in \ell^2(\mathbb{Z}_N)$, we define an inner product on $\ell^2(\mathbb{Z}_n)$ by

$$\langle u, v \rangle = \sum_{k=0}^{N-1} u(k)\overline{v(k)}.$$

It is straightforward to show that this is an inner product; from a computational perspective, this inner product is identical to the dot product on $\mathbb{C}^N$.

The DFT projects an arbitrary signal $z \in \ell^2(\mathbb{Z}_N)$ onto an orthonormal basis; again, the process is identical to the process of projecting a vector onto a basis that we introduced in Chapter 1. We now introduce the set of orthonormal basis vectors for $\ell^2(\mathbb{Z}_N)$ that we will need for the DFT.[1]

[1]As we will soon see, we will ultimately alter the weights on these basis vectors slightly so that the set is orthogonal but not orthonormal.

DEFINITION 20. We define $E_0, E_1, ..., E_{N-1} \in \ell^2(\mathbb{Z}_N)$ as

$$\begin{aligned}
E_0 &= \frac{1}{\sqrt{N}} && \textit{for } n = 0, 1, ..., N-1;\\
E_1 &= \frac{1}{\sqrt{N}} e^{2\pi i n/N} && \textit{for } n = 0, 1, ..., N-1;\\
E_2 &= \frac{1}{\sqrt{N}} e^{2\pi i 2n/N} && \textit{for } n = 0, 1, ..., N-1;\\
&\vdots\\
E_{N-1} &= \frac{1}{\sqrt{N}} e^{2\pi i (N-1)n/N} && \textit{for } n = 0, 1, ..., N-1.
\end{aligned}$$

THEOREM 2.2. The set $\{E_0, ..., E_{N-1}\}$ is an orthonormal basis for $\ell^2(\mathbb{Z}_N)$.

PROOF. We first need to show that $\{E_0, ..., E_{N-1}\}$ is an orthogonal set. Suppose $j, k \in \{0, 1, ..., N-1\}$. Then

$$\begin{aligned}
\langle E_j, E_k \rangle &= \sum_{n=0}^{N-1} E_j(n) \overline{E_k(n)}\\
&= \sum_{n=0}^{N-1} \frac{1}{\sqrt{N}} e^{2\pi i j n/N} \overline{\frac{1}{\sqrt{N}} e^{2\pi i k n/N}}\\
&= \frac{1}{N} \sum_{n=0}^{N-1} e^{2\pi i j n/N} e^{-2\pi i k n/N}\\
&= \frac{1}{N} \sum_{n=0}^{N-1} e^{2\pi i (j-k)n/N}\\
&= \frac{1}{N} \sum_{n=0}^{N-1} \left(e^{(2\pi i (j-k)/N)}\right)^n .
\end{aligned}$$

If $j = k$, then $\langle E_j, E_k \rangle = \frac{1}{N} \sum_{n=0}^{N-1} 1 = 1$. Therefore, $||E_j|| = 1$ for all j. If $j \neq k$, then

$$\frac{1}{N} \sum_{n=0}^{N-1} \left(e^{(2\pi i (j-k)/N)}\right)^n = \frac{1 - \left(e^{2\pi i (j-k)/N}\right)^N}{1 - e^{2\pi i (j-k)/N}}$$

since the sum is a partial sum of a geometric series. Then

$$(e^{2\pi i(j-k)/N})^N = e^{2\pi i(j-k)} = 1$$

since $j, k \in \mathbb{Z}$. Thus, $\langle E_j, E_k \rangle = 0$ for $j \neq k$. Therefore, $\{E_0, ..., E_{N-1}\}$ is an orthonormal set. From Theorem 1.2, we know that an orthonormal set is linearly independent. We know that $\dim \ell^2(\mathbb{Z}_N) = \dim\{E_0, ..., E_{N-1}\} = N$. Then (though we did not mention this fact in Chapter 1) since the dimension of the basis is the same as the dimension of the vector space, $\{E_0, ..., E_{N-1}\}$ is an orthonormal basis for $\ell^2(\mathbb{Z}_N)$. □

From Euler's formula, we know that

$$e^{i\theta} = \cos\theta + i\sin\theta.$$

Then for an arbitrary m,

$$E_m(n) = \cos\left(\frac{2\pi mn}{N}\right) + i\sin\left(\frac{2\pi mn}{N}\right).$$

That is, $E_m(n)$ is the sum of two sinusoids of frequency m/N that are evenly sampled at values from 0 to $\frac{N-1}{N}$.

We now define the Discrete Fourier transform (DFT) and the Inverse Discrete Fourier transform (IDFT)[2].

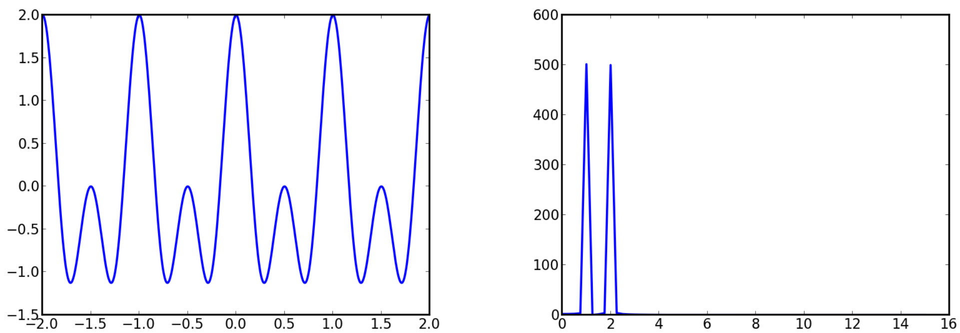
Definition 21. Suppose $z = (z(0), ..., z(N-1)) \in \ell^2(\mathbb{Z}_N)$. For $m = 0, 1, ..., N-1$, define

$$\mathcal{F}(z)(m) = \sum_{n=0}^{N-1} z(n)e^{-2\pi imn/N}. \tag{2.1}$$

Let $\mathcal{F}(z) = (\mathcal{F}(z)(0), \mathcal{F}(z)(1), ..., \mathcal{F}(z)(N-1))$; therefore, $\mathcal{F}(z) \in \ell^2(\mathbb{Z}_N)$. This mapping from $\ell^2(\mathbb{Z}_N) \to \ell^2(\mathbb{Z}_N)$ is called the **Discrete Fourier transform**.

We should identify the linear algebra concepts that led us to this point; in particular, note that

[2]For computational reasons, we don't actually compute the DFT with the set $\{E_0, ..., E_{N-1}\}$: we drop the $\frac{1}{N}$ multiplier.



(a) Time-Amplitude Plot (b) Frequency-Amplitude Plot

FIGURE 2.1. Plots of $\cos(2\pi x) + \cos(4\pi x)$

$$\mathcal{F}(z)(m) = \sqrt{N}\,\langle z, E_m \rangle$$

and

$$\mathcal{F}(z) = \sqrt{N}[z]_{E_m}.$$

DEFINITION 22. Let $w = (w(0), ..., w(N-1)) \in \ell^2(\mathbb{Z}_N)$. Then for $n = 0, 1, ..., N-1$, define

$$\mathcal{F}^{-1}(w)(n) = \frac{1}{N}\sum_{m=0}^{N-1} w(m)e^{2\pi i m n/N}. \tag{2.2}$$

Then $\mathcal{F}^{-1}(w) = (\mathcal{F}^{-1}(w)(0), \mathcal{F}^{-1}(w)(1), ..., \mathcal{F}^{-1}(w)(N-1))$. This mapping from $\ell^2(\mathbb{Z}_N) \to \ell^2(\mathbb{Z}_N)$ is called the **Inverse Discrete Fourier transform**.

THEOREM 2.3. Let $z \in \ell^2(\mathbb{Z}_N)$. Then $\mathcal{F}^{-1}(\mathcal{F}(z))(n) = z(n)$.

PROOF. By definition and using Theorem 1.1,

$$\begin{aligned}
\mathcal{F}^{-1}(\mathcal{F}(z))(k) &= \frac{1}{N}\sum_{m=0}^{N-1} \mathcal{F}(z)(m) \cdot e^{2\pi imk/N} \\
&= \frac{1}{N}\sum_{m=0}^{N-1} \left(\sqrt{N}\,\langle z, E_m\rangle\right) e^{2\pi imk/N} \\
&= \frac{1}{N}\sum_{m=0}^{N-1} \left(\sqrt{N}\,\langle z, E_m\rangle\right)\left(\sqrt{N}E_m(k)\right) \\
&= \sum_{m=0}^{N-1} \langle z, E_m\rangle\, E_m(k) \\
&= z(k).
\end{aligned}$$

□

The IDFT allows us to reconstruct a signal from its Fourier transform. However, we can also use the IDFT to construct a signal with the frequency content that we want; we will use this property in Chapter 3 to construct wavelets.

As a terminological note, for $w \in \ell^2(\mathbb{Z}_N)$, we call say $\mathcal{F}(w)$ is a function in the **frequency domain**; in fact, we will refer to any function in $\ell^2(\mathbb{Z}_N)$ whose values we interpret as magnitudes of frequencies as a function in the frequency domain. Similarly, for $z \in \ell^2(\mathbb{Z}_N)$, we say that $\mathcal{F}^{-1}(z)$ is a function in the **time domain**; we will refer to any function in $\ell^2(\mathbb{Z}_N)$ whose values we interpret as magnitudes of a signal at a given time as a function in the time domain. Of course, both w and z are functions in $\ell^2(\mathbb{Z}_N)$; the real difference between the frequency and the time domains is the way that we interpret functions. In later chapters, we will construct signals in the frequency domain; when we take the IDFT of these signals, we will have a signal in the time-domain with the desired frequency properties.

1.1. Resolution and the Basic Properties of the DFT. We will now discuss spatial and frequency **resolution** of bases of $\ell^2(\mathbb{Z}_N)$. Consider the *standard* or *Euclidean* basis for $\ell^2(\mathbb{Z}_N)$ that we introduced earlier. If we project a signal onto the standard basis, then - since the projection will just be the original signal - we will know *all* of the spatial properties of our original signal: we know the exact value of $z(n)$ for all n.

DEFINITION 23. For any $z \in \ell^2(\mathbb{Z}_N)$, define the mapping from $\mathcal{I} : \ell^2(\mathbb{Z}_N) \to \ell^2(\mathbb{Z}_N)$ given by $\mathcal{I}(z) = z$ the **standard transform**. We say that $\mathcal{I}(z)$ has **perfect spatial resolution on** z.

The Fourier transform projects a signal onto pure sinusoids. Since sinusoids are the basic unit of frequency, it makes sense to *define* the frequency content of a signal as the Fourier transform of a signal.

DEFINITION 24. For any $z \in \ell^2(\mathbb{Z}_N)$, we say that $\mathcal{F}(z)$ has **perfect frequency resolution on** z.

For a given z, we should expect that the standard basis will tell us nothing about the frequency content of z; we can prove this mathematically.

THEOREM 2.4. Let $\{\mathbf{e}_j(n)\}_{j=0}^{N-1}$ be the standard basis for $\ell^2(\mathbb{Z}_N)$. Then for all j,

$$|\mathcal{F}(\mathbf{e}_j)(m)| = 1 \text{ for } m = 0, 1, ..., N-1.$$

PROOF. By definition,

$$\begin{aligned}\mathcal{F}(\mathbf{e}_i)(m) &= \sum_{n=0}^{N-1} \mathbf{e}_j(n) e^{-2\pi imn/N} \\ &= e^{-2\pi imj/N}.\end{aligned}$$

Since $|e^{-2\pi imj/N}| = 1$, this is the desired result.

□

That is, each vector in the standard basis "contains" all frequencies evenly.

We now want to examine the spatial resolution of the Fourier transform. We will do this by examining how the Fourier transform responds to a shifted signal. We will show that the Fourier transform can only reveal phasing[3] information about a signal – we will examine how the Fourier transform responds to a shifted signal.

DEFINITION 25. Suppose $z \in \ell^2(\mathbb{Z}_N)$ and $k \in \mathbb{Z}$. Define

$$(R_k z)(n) = z((n-k) \bmod N))$$

for $n \in \mathbb{Z}$.

EXAMPLE 13.

Suppose

$$z = [1, 2, 3, 4].$$

Then

$$R_2 z = [3, 4, 1, 2].$$

THEOREM 2.5. Suppose $z \in \ell^2(\mathbb{Z})$ and $k \in \mathbb{Z}$. Then for any $m \in \mathbb{Z}$,

$$\mathcal{F}(R_k z)(m) = e^{-2\pi imn/N} \mathcal{F}(z)(m).$$

Then

$$|\mathcal{F}(R_k z)(m)| = |\mathcal{F}(z)(m)|.$$

That is, the Fourier transform is **translation-invariant**.

PROOF. By definition,

$$\mathcal{F}(R_k z)(m) = \sum_{n=0}^{N-1} (R_k z)(n) e^{-2\pi imn/N} = \sum_{n=0}^{N-1} z(n-k) e^{-2\pi imn/N}.$$

We now change variables by letting $\ell = n - k$. Then we have

$$\mathcal{F}(R_k z)(m) = \sum_{\ell=-k}^{N-k-1} z(\ell) e^{-2\pi im(\ell+k)/N} = e^{-2\pi imk/N} \sum_{\ell=-k}^{N-k-1} z(\ell) e^{-2\pi im(\ell)/N}.$$

[3]The word "phase" actually does refer to the location of a signal, but only with respect to a sine or cosine function.

We must now show that

$$\sum_{\ell=-k}^{N-k-1} z(\ell)e^{-2\pi i m(\ell)/N} = \mathcal{F}(z)(m).$$

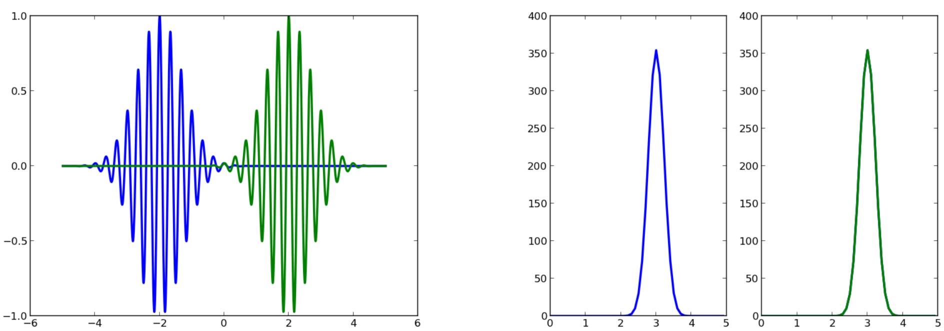
Remember that $z(\ell)$ and $e^{-2\pi i m\ell/N}$ are periodic functions with period N. Then split up the summation:

$$\sum_{\ell=-k}^{N-k-1} z(\ell)e^{-2\pi i m(\ell)/N} = \sum_{\ell=-k}^{-1} z(\ell+N)e^{-2\pi i m(\ell+N)/N} + \sum_{\ell=0}^{N-k-1} (z)\ell e^{-2\pi i m\ell/N}.$$

If we let $n = \ell + N$ in the first sum and $n = \ell$ in the second sum, then

$$\begin{aligned}\sum_{\ell=-k}^{N-k-1} z(\ell)e^{-2\pi i m(\ell)/N} &= \sum_{n=N-k}^{N-1} z(n)e^{-2\pi i mn/N} + \sum_{n=0}^{N-k-1} z(n)e^{-2\pi i mn/N} \\ &= \sum_{n=0}^{N-1} z(n)e^{-2\pi i mn/N}.\end{aligned}$$

□



(a) Time-Amplitude Plot (b) Frequency-Amplitude Plots

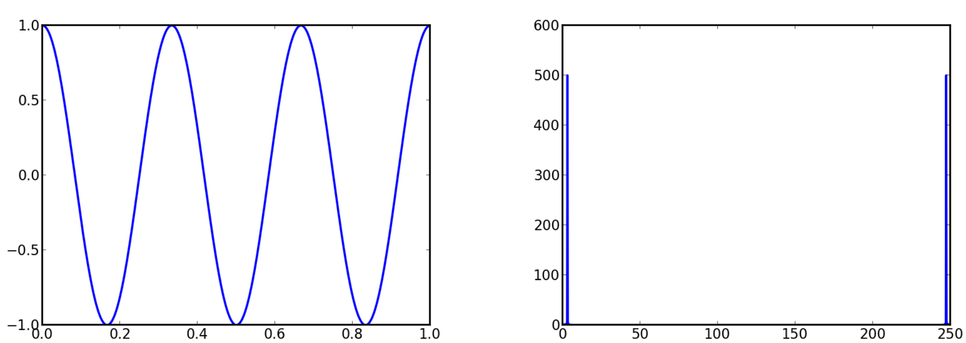
FIGURE 2.2. Time Shifted Signals

It's hard to overemphasize the importance of this result: the DFT cannot differentiate a signal from a shifted version of itself - see Figure 2.2. In many ways, this is a good thing: the frequency content of a signal should not depend upon its phase. However, this property also clearly limits our ability to analyze the local properties of a signal.

It turns out that there are fundamental limits on our ability to mathematically spatially locate a given frequency - this limit is actually given by the Heisenberg uncertainty principle, a well known result from quantum mechanics. Though we will not discuss the uncertainty principle in detail, we will roughly state its main result: the more spatial information we know, the less frequency information we can know, and vice versa. Conceptually, the wavelet transform allows us to navigate this limit stipulated by the Heisenberg uncertainty principle by allowing us to know *some* frequency information and *some* time information simultaneously. As I mentioned earlier, it turns out that the wavelet transform is a projection onto a new orthogonal basis. Unlike the standard basis - which has no frequency resolution - and the Fourier basis - which has no spatial resolution - the wavelet basis will have some frequency and spatial resolution simultaneously.

1.2. A Brief Note on Aliasing. We need to consider a technical problem that we have completely avoided up until this point. In discrete signal processing, **aliasing** arises as a result of the discrete sampling of sinusoids; when we sample a high-frequency signal too infrequently, our sampled signal will *look* like a low-frequency signal.

This problem arises when we examine the Discrete Fourier transform. All of the graphs labeled "Frequency-Amplitude Plots" in this thesis have been truncated plots; I have only plotted half of the values. Figure 2.3 shows a full plot of the DFT of $\cos(6\pi x)$ plotted on $[0, 1]$ and sampled 1000 times:



(a) T-A Plot (b) F-A Plot

Figure 2.3. Full DFT of $\cos(6\pi x)$

A full plot of the DFT will always be symmetric about the midpoint of the horizontal axis. Why does this happen? Consider the following two sinusoids at frequencies $\frac{0}{N}$ and $\frac{N}{N}$; the projection of a signal onto these sinusoids should be plotted at the opposite ends of the DFT:

$$\sin\left(\frac{2\pi(0)n}{N}\right) = 0$$

and

$$\sin\left(\frac{2\pi(N)n}{N}\right) = \sin\left(2\pi n\right).$$

Considered as continuous functions—i.e., $n \in \mathbb{R}$—these two sinusoids look totally different. However, when $n \in \mathbb{N}$, these signals are indistinguishable. And this problem exists across the frequency spectrum: a $\frac{1}{N}$ frequency signal will look exactly like a $\frac{N-1}{N}$ frequency signal when sampled discretely, a $\frac{2}{N}$ frequency signal will look like a $\frac{N-2}{N}$ frequency signal, and so on.

From now on, we say that a given sampling rate can only **support** a certain range of frequencies. In fact, the **Nyquist frequency**, named after Harry Nyquist, of a signal is defined as one-half of the sampling frequency of that signal[4] and is the maximum frequency it can support.

In summary, we mention aliasing because it will affect many of our later constructions—in the frequency domain, a low-frequency signal will actually appear to contain low and very high frequencies.

1.3. Convolution. Most high-quality speaker systems have knobs or sliders with the labels "*low*", "*mid*", and "*high*". As users adjust these knobs, the speaker system augments or attenuates the low, middle, or high-range frequencies of the music playing through the system.

[4]Fun fact: the average human being can only perceive sound up to about 22 MHz; the conventional audio sampling rate is about 44 MHz. Coincidence?

Electrical engineers design the filters[5] contained within speaker systems; the action of these electronic filters on an input signal can be described mathematically by the convolution operator. Though a speaker system is actually a continuous system, the same principles of filtration apply to discrete systems[6]. Though it will not be clear now, it turns out that the discrete wavelet transform can be characterized as an iterative filtration problem.

In this chapter, we will introduce the convolution operator and some of its important theorems; as a highlight of the chapter, we will reveal the intimate relationship between the Fourier transform and convolution.

DEFINITION 26. Let $z, w \in \ell^2(\mathbb{Z}_N)$. For $m = 0, 1, ..., N-1$, define

$$(z * w)(m) = \sum_{n=0}^{N-1} z(m-n)w(n). \tag{2.3}$$

Let $z * w = ((z*w)(0), (z*w)(1), ..., (z*w)(N-1))$; then $(z*w) \in \ell^2(\mathbb{Z}_N)$. This mapping from $\ell^2(\mathbb{Z}_N) \to \ell^2(\mathbb{Z}_N)$ is called the **convolution** of z with w.

EXAMPLE 14.

Let $z, w \in \ell^2(\mathbb{Z}_3)$, where $z = [1, 2, 3]$ and $w = [1, 1, 1]$. Then

$$(z*w)(0) = z(0-0)w(0) + z(0-1)w(1) + z(0-2)w(2) = 1\cdot 1 + 3\cdot 1 + 2\cdot 1 = 6$$

$$(z*w)(1) = z(1-0)w(0) + z(1-1)w(1) + z(1-2)w(2) = 2\cdot 1 + 1\cdot 1 + 3\cdot 1 = 6$$

$$(z*w)(2) = z(2-0)w(0) + z(2-1)w(1) + z(2-2)w(2) = 3\cdot 1 + 2\cdot 1 + 1\cdot 1 = 6$$

Then $z * w = [6, 6, 6]$.

[5]By filtration, we *only* mean the attenuation or augmentation of frequencies. The field of signal processing considers all kinds of other "filters" like delays or phasers—think of the effects produced by a guitar pedal. These "filters", while interesting, aren't related to convolution, and they aren't relevant to this thesis.

[6]A linear time-invariant system, like the filters within speaker system (and the speaker system itself), is characterized by its **impulse response**. The response of a filter to a signal is given by the convolution of the signal with the filter's impulse response.

With a simple change of variables, it can be proved that convolution is a commutative operator; that is,

$$z * w = w * z.$$

1.4. Properties of the Convolution Operator. The definition of convolution doesn't tell us much about how z filters w or vice versa; it's not a particularly intuitive definition[7]. However, the following theorem—often called "The Convolution Theorem"[8]—will reveal how convolution acts on the frequency content of z and w. The convolution theorem will be a central theorem in this thesis; it demonstrates the fundamental relationship between convolution and the Fourier transform.

THEOREM 2.6. **The Convolution Theorem.** Suppose $z, w \in \ell^2(\mathbb{Z}_N)$. Then for each m,

$$\mathcal{F}(z * w)(m) = \mathcal{F}(z)(m)\mathcal{F}(w)(m) \tag{2.4}$$

or

$$\mathcal{F}(z * w) = \mathcal{F}(z)\mathcal{F}(w). \tag{2.5}$$

[7]Sometimes texts will say that the convolution operator is a "reverse, shift, and sum" operator—a graphical description of what the operator is doing. There are plenty of diagrams online that demonstrate this process; they give some intuition about how the convolved signal $z*w$ is a mix of both z and w. However, they don't give too many hints about relationship between the frequency content of $z*w, z$, and w - this is what we're really interested in.

[8]The theorem below will be numbered like the rest of the theorems in this thesis. However, I will continue to refer to it as "the convolution theorem" because it will make the text easier to read.

PROOF. By definition,

$$\begin{aligned}
\mathcal{F}(z*w)(m) &= \sum_{n=0}^{N-1}(z*w)(n)e^{-2\pi imn/N} \\
&= \sum_{n=0}^{N-1}\sum_{k=0}^{N-1} z(n-k)w(k)e^{-2\pi imn/N} \\
&= \sum_{n=0}^{N-1}\sum_{k=0}^{N-1} z(n-k)w(k)e^{-2\pi im(n-k)/N}e^{-2\pi imk/N} \\
&= \left(\sum_{k=0}^{N-1} w(k)e^{-2\pi imk/N}\right)\left(\sum_{n=0}^{N-1} z(n-k)e^{-2\pi im(n-k)/N}\right).
\end{aligned}$$

In the second sum, we change index to let $\ell = n-k$ so that

$$\begin{aligned}
\sum_{n=0}^{N-1} z(n-k)e^{-2\pi im(n-k)/N} &= \sum_{\ell=-k}^{N-1-k} z(\ell)e^{-2\pi im\ell/N} \\
&= \sum_{\ell=0}^{N-1} z(\ell)e^{-2\pi im\ell/N}
\end{aligned}$$

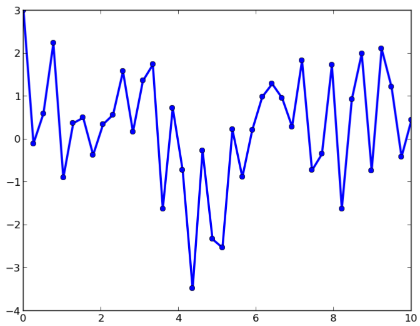
since $z(\ell)$ and $e^{-2\pi im\ell/N}$ are N-periodic. Then

$$\begin{aligned}
\mathcal{F}(z*w)(m) &= \left(\sum_{k=0}^{N-1} w(k)e^{-2\pi imk/N}\right)\left(\sum_{\ell=0}^{N-1} z(\ell)e^{-2\pi im\ell/N}\right) \\
&= \mathcal{F}(z)(m)\mathcal{F}(w)(m).
\end{aligned}$$

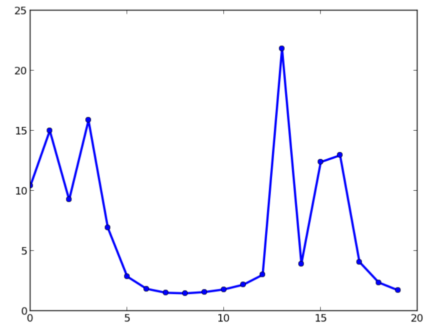
□

In words, the convolution theorem states that the frequency content of two convolved functions equals the product of the frequency content of the original functions; examine Figure 2.4 carefully to see how this process works out.

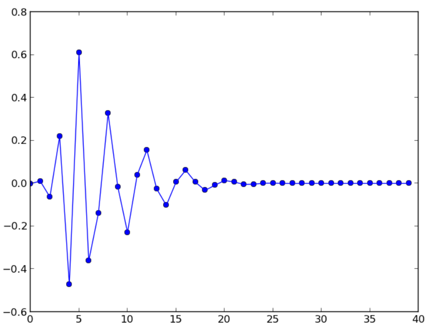
The convolution theorem also reveals how we can construct a filter to augment or attenuate frequencies in another signal: if we construct a function in the frequency domain with the desired frequency properties, the IDFT of our function will have the desired frequency



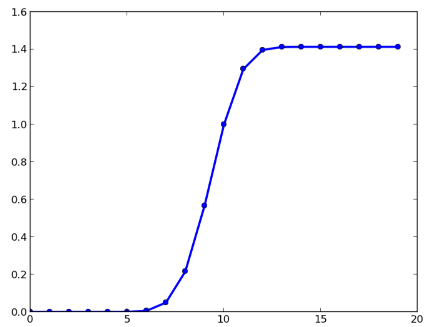
(a) T-A Plot of Sample Signal



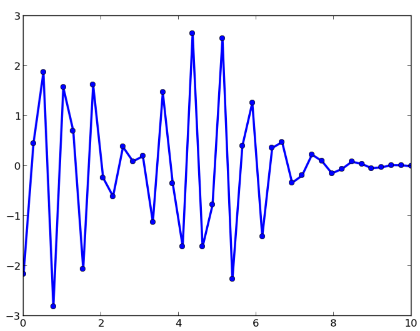
(b) F-A Plot of Sample Signal



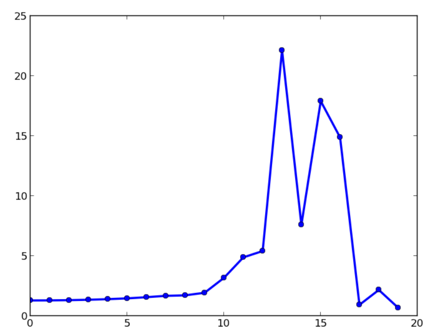
(c) T-A Plot of High-Pass Filter



(d) F-A Plot of High-Pass Filter



(e) T-A Plot of Signal After Convolution



(f) F-A Plot of Signal After Convolution

FIGURE 2.4. Convolution of Sample Signal with High-Pass Filter

properties. When convolved with a signal, our constructed function will filter that signal in accordance with the shape of its frequency-domain.

We have names for particular kinds of filters:

- A **high-pass** filter will attenuate low frequencies and leave high frequencies unchanged.
- A **low-pass** filter will attenuate high frequencies and leave low frequencies unchanged.

Depending upon the functional shape of the filter in the frequency domain, a high- or low-pass filter may be imperfect[9]—it may filter frequencies unevenly or allow some unwanted frequencies.

2. Convolution and Projection

It turns out that convolution can also be interpreted as the projection of a signal onto a certain kind of basis. For $w \in \ell^2(\mathbb{Z}_N)$, consider the set $W = \{R_k w\}_{k=0}^{N-1}$; we say that w is the **generator** of W and that W is the **set generated by** w. Depending on what w is, W can be a basis for $\ell^2(\mathbb{Z}_N)$.

EXAMPLE 15. Consider $\mathbf{e}_0$, the first vector in the standard basis of $\ell^2(\mathbb{Z}_N)$. Then the set generated by $\mathbf{e}_0$, $\{R_k\mathbf{e}_0\}_{k=0}^{N-1}$, will be the entire standard basis.

We will now show that convolution can project a signal onto a basis of this form.

DEFINITION 27. For any $w \in \ell^2(\mathbb{Z}_N)$, define $\tilde{w} \in \ell^2(\mathbb{Z}_N)$ by

$$\tilde{w}(n) = \overline{w(-n)} = \overline{w(N-n)}.$$

We call $\tilde{w}$ the **conjugate reflection** of w.

As we should expect, the Fourier transform of the conjugate reflection of a signal is closely related to the original signal; the following theorem will be important for future proofs.

THEOREM 2.7. Let $z \in \ell^2(\mathbb{Z}_N)$. Then

$$\mathcal{F}(\tilde{z}) = \overline{\mathcal{F}(z)}.$$

[9]Continuous filters are *always* imperfect.

PROOF. Let $z \in \ell^2(\mathbb{Z}_N)$. First note that $z(-n) = z(N-n)$. Then

$$\begin{aligned}
\mathcal{F}(\tilde{z})(k) &= \sum_{k=0}^{N-1} \tilde{z} e^{-2\pi ikn/N} \\
&= \sum_{k=0}^{N-1} \overline{z(N-n)} e^{-2\pi ikn/N} \\
&= \overline{\sum_{k=0}^{N-1} z(N-n) e^{-2\pi ik(N-n)/N}} \\
&= \overline{\mathcal{F}(z)(k)}
\end{aligned}$$

□

Equally important, however, is the following theorem.

THEOREM 2.8. Suppose $z, w \in \ell^2(\mathbb{Z}_N)$. For any $k \in \mathbb{Z}$,

$$(z * \tilde{w})(k) = \langle z, R_k w \rangle \tag{2.6}$$

PROOF. By definition,

$$\begin{aligned}
\langle z, R_k w \rangle &= \sum_{n=0}^{N-1} z(n) \overline{R_k w(n)} \\
&= \sum_{n=0}^{N-1} z(n) \tilde{w}(k-n) \\
&= (\tilde{w} * z)(k) \\
&= (z * \tilde{w})(k)
\end{aligned}$$

□

Furthermore, suppose that $w \in \ell^2(\mathbb{Z}_N)$ and $W = \{R_k w\}_{k=0}^{N-1}$ is an orthonormal basis for $\ell^2(\mathbb{Z}_N)$. Then,

$$z * \tilde{w} = [z]_W. \tag{2.7}$$

In words, the convolution of the conjugate reflection of a function w and another function z is the projection of z onto the basis generated by w. Looking ahead to the next chapter, it turns out that the wavelet basis will be a basis of exactly this form; we will compute the wavelet transform of a function by convolving it with a wavelet basis vector.

2.1. Power Spectrum Estimation. In general, we want to be more careful about how we interpret the information given to us by the Fourier transform. Though the raw Fourier transform is useful for many applications - especially those that require an orthogonal basis for $\ell^2(\mathbb{Z}_N)$ - it does not always accurately reflect the frequency content of an underlying signal; we will clarify what we mean by this in the upcoming discussion.

Before we begin, however, we want to introduce two theorems that, together, mathematically demonstrate the relationship between the *power* of a signal and its Fourier transform.

DEFINITION 28. Let $z \in \ell^2(\mathbb{Z}_N)$. Then the **power** of z is given by

$$||z||^2 = \langle z, z \rangle .$$

The power of a signal (or of a Fourier transformed signal) consolidates the information given by its real and complex components.

THEOREM 2.9. **Parseval's Relation**. Let $z, w \in \ell^2(\mathbb{Z}_N)$. Then

$$\langle z, w \rangle = \frac{1}{N} \langle \mathcal{F}(z), \mathcal{F}(w) \rangle .$$

PROOF. First, we know from Theorem 1.1 that

$$z = \sum_{i=0}^{N-1} \langle z, E_i \rangle E_i$$

and

$$w = \sum_{i=0}^{N-1} \langle w, E_i \rangle E_i$$

since $\{E_i\}_{i=0}^{N-1}$ is an orthonormal basis. Then

$$\begin{aligned}
\langle z, w \rangle &= \left\langle \sum_{i=0}^{N-1} \langle z, E_i \rangle E_i, \sum_{j=0}^{N-1} \langle w, E_j \rangle E_j \right\rangle \\
&= \sum_{i=0}^{N-1} \langle z, E_i \rangle \sum_{j=0}^{N-1} \langle E_i, \langle w, E_j \rangle E_j \rangle \\
&= \sum_{i=0}^{N-1} \langle z, E_i \rangle \sum_{j=0}^{N-1} \overline{\langle w, E_j \rangle} \langle E_i, E_j \rangle
\end{aligned}$$

using the linearity properties of inner products. Since

$$\langle E_i, E_j \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

we see that $\langle z, w \rangle = \sum_{i=0}^{N-1} \langle z, E_i \rangle \overline{\langle w, E_j \rangle}$. But we know that $\mathcal{F}(z)(m) = \sqrt{N} \langle z, E_m \rangle$, therefore

$$\begin{aligned}
\langle z, w \rangle &= \sum_{i=0}^{N-1} \frac{1}{\sqrt{N}} \mathcal{F}(z)(i) \overline{\frac{1}{\sqrt{N}} \mathcal{F}(w)(i)} \\
&= \frac{1}{N} \langle \mathcal{F}(z), \mathcal{F}(w) \rangle .
\end{aligned}$$

□

Note that the validity of Parserval's relation relies *only* on the orthogonality of the Fourier basis – very similar theorems will hold true for other orthogonal and orthonormal bases, including wavelet bases.

THEOREM 2.10. **Plancherel's formula**. Let $z, w \in \ell^2(\mathbb{Z}_N)$. Then

$$||z||^2 = \frac{1}{N}||\mathcal{F}(z)||^2.$$

PROOF. Plancherel's formula follows straightforwardly from Parseval's relation by letting $w = z$. □

Let's examine Plancherel's formula more carefully. First, it states that the power of a signal z is directly proportional to the power of its Fourier transform. Secondly, we can re-write $||\mathcal{F}(z)||^2$ as a sum; that is,

$$||\mathcal{F}(z)||^2 = \sum_{i=0}^{N-1} |\mathcal{F}(z)(i)|^2.$$

By examining $|\mathcal{F}(z)(i)|^2$ for each i, it *seems* that we can immediately evaluate how different frequencies[10] contribute to the overall power of the original signal z. Unfortunately, as we will see in the following discussion, the analysis of signal power spectra is a more messy process than we might expect - see [**1**].

Let $\omega_0, \omega_1, ..., \omega_{N-1}$ be the frequencies measured "exactly" by the Fourier transform; that is, from the definition of the Fourier transform, $\omega_j = \frac{2\pi j}{N}$ for some $j \in \{0, 1, ..., N-1\}$. For each ω_j, define

$$y_j(n) = \cos(\omega_j n).$$

Then $|\mathcal{F}(y_j)(i)|^2$ will have a non-zero value *only* at $m = j$ (and, because of aliasing, at $i = N - j$).

However, if we consider

$$y(n) = \cos(\omega n)$$

[10]The terms $|\mathcal{F}(z)(i)|^2$ are often plotted; these plots are called **periodograms**.

where $\omega \neq \omega_j$ for all j, then, perhaps surprisingly, $|\mathcal{F}(y)(i)|^2$ will be non-zero for all i. In fact, for all j, the power of y will be diffused across Fourier coefficients a_j and b_j at a rate proportional to $1/|\omega - \omega_j|$; to re-emphasize, *all* of the Fourier coefficients will be affected by a signal containing a frequency not measured exactly by the Fourier transform.

We call this phenomenon **spectral leakage**. For many applications, it is undesirable, especially when we are trying to identify the frequency components of a signal.

It turns out that we can reduce this spectral leakage by filtering the Fourier transform coefficients a_i and b_i. If we denote our filtered Fourier coefficients by A_i and B_i, then

$$\begin{aligned} A_i &= -\frac{1}{4}a_{i-1} + \frac{1}{2}a_i - \frac{1}{4}a_{i+1} \\ B_i &= -\frac{1}{4}b_{i-1} + \frac{1}{2}b_i - \frac{1}{4}b_{i+1}. \end{aligned}$$

This kind of smoothing can be achieved by convolving the real and imaginary parts of the Fourier transform with an appropriately designed filter.

Definition 29. We define the **Hanning filter** $H_f \in \ell^2(\mathbb{Z}_N)$ for $z \in \ell^2(\mathbb{Z}_N)$ as follows.

$$H_f(i) = \begin{cases} -\frac{1}{4} & i = 0 \\ \frac{1}{2} & i = 1 \\ -\frac{1}{4} & i = 2 \\ 0 & \text{elsewhere.} \end{cases}$$

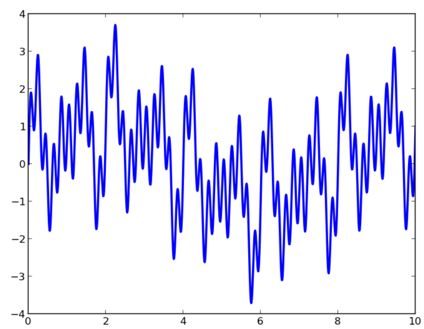
Definition 30. Let $z \in \ell^2(\mathbb{Z}_N)$. Then for $i = 0, 1, ..., N-1$, the Hanning filtered Fourier coefficients of $\mathcal{F}[z]$ are given by

$$H_f[\mathcal{F}[z]] = \mathcal{F}[z] * H_f \tag{2.8}$$

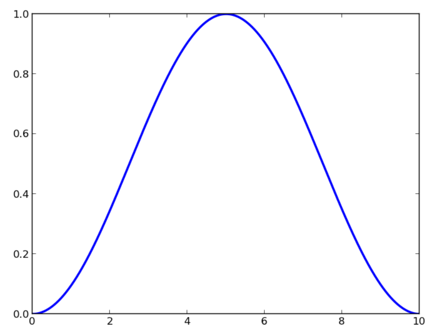
or, alternatively, using the convolution theorem,

$$H_f[\mathcal{F}[z]] = \mathcal{F}^{-1}\left[\mathcal{F}\left[\mathcal{F}[z]\right]\mathcal{F}[H_f]\right]. \tag{2.9}$$

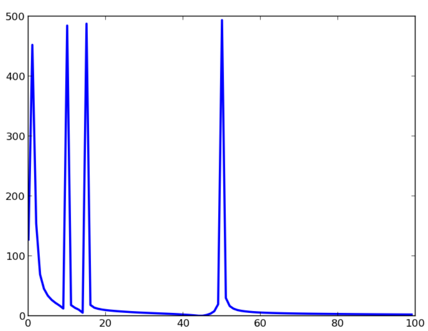
The Python code for implementing the Hanning filter can be found in the Appendix.



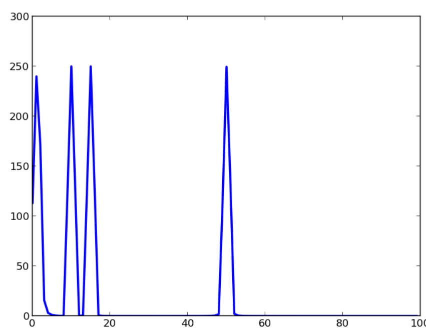
(a) $f(x) = \sin(2\pi x) + \sin(3\pi x) + \sin(10\pi x) + \sin(\frac{1}{4}\pi x)$



(b) Fourier transform of Hanning filter



(c) FFT of $f(x)$ without Hanning filter



(d) FFT of $f(x)$ with Hanning filter

FIGURE 2.5

Consider Figure 2.5. In particular, note the non-zero coefficients on either side of the peak frequencies in Figure 2.5.(c).—this is the "spectral leakage"— and their absence in Figure 2.5.(d).

Why does Hanning filtration work? Figure 2.5.b. shows the Fourier transform of the Hanning filter - note that the filter peaks at the highest frequencies (again, remember aliasing) . Importantly, this means that the Hanning filter will only let sharp peaks in the Fourier coefficients pass; the non-zero coefficients around the peak frequencies are essentially low frequency signals and are therefore attenuated by the Hanning filter.

CHAPTER 3

The Discrete Wavelet Transform

1. The Haar Wavelet

I hope that I have presented the material up until this point in a way that the upcoming discussion will make the goals of and motivation for the wavelet transform fairly intuitive. Nevertheless, I want to provide an example of the phenomenon introduced in Theorem 2.5—that the Fourier transform does not respond to shifted signals. It will emphasize the aims of the whole thesis and serve as a good mathematical transition to wavelets.

EXAMPLE 16. Let $N = 4$. Then consider the set $\{E_k\}_{k=0}^{3}$.

$$\begin{aligned} E_0 &= \frac{1}{2}(1, 1, 1, 1) \\ E_1 &= \frac{1}{2}(1, i, -1, -i) \\ E_2 &= \frac{1}{2}(1, -1, 1, -1) \\ E_3 &= \frac{1}{2}(1, -i, 1, -i) \end{aligned}$$

Visually, it's clear that these basis vectors have frequency properties; Chapter 2 articulated these properties and the DFT's ability to exploit them. Yet the DFT is unable to differentiate signals with localized time properties. Consider the signals $u_1, u_2 \in \ell^2(\mathbb{Z}_4)$ that represent simple localized oscillations:

$$u_1 = \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, 0, 0\right)$$
$$u_2 = \left(0, 0, \frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)$$

The DFT of u_1, u_2 should be similar[1]:

$$\mathcal{F}(u_1) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2} \\ \sqrt{2} \\ \frac{\sqrt{2}}{2} - i\frac{\sqrt{2}}{2} \end{pmatrix}$$

$$\mathcal{F}(u_2) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix} = \begin{pmatrix} 0 \\ -\frac{\sqrt{2}}{2} - i\frac{\sqrt{2}}{2} \\ \sqrt{2} \\ -\frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2} \end{pmatrix}$$

In fact, as we proved in Chapter 2, the moduli of $\mathcal{F}(u_1)$ and $\mathcal{F}(u_2)$ should be identical:

$$|\mathcal{F}(u_1)| = |\mathcal{F}(u_2)| = \begin{pmatrix} 0 \\ 1 \\ \sqrt{2} \\ 1 \end{pmatrix}$$

Again, this result is not unexpected: the frequency content of a signal should not depend on the spatial location of the signal. Nevertheless, we lose all of the time information contained within a signal when we compute its DFT—this information is important for

[1]We compute the Fourier transform of u_1 and u_2 as a matrix-vector product; it is equivalent to the DFT presented earlier. Note, however, that your computer does not compute the Fourier transform this way: developed by Tukey and Cooley in 1965, the Fast Fourier transform (FFT) reduces the computation time of the Fourier transform from $O(n^2)$—the usual computation time for a matrix-vector product— to $O(n \log n)$ operations.

analyzing transients. The central insight of the wavelet transform is that the best way to analyze transients is to construct an orthogonal basis of *transients* with some wave-like properties—rather than a basis of pure sinusoids. In fact, the first wavelet basis—the Haar wavelet basis—is constructed in part by using u_1 and u_2 as basis vectors.

1.1. The Haar Wavelet transform. Like before, let's only consider $\ell^2(\mathbb{Z}_4)$ for the moment. Introduced by Alfred Haar in 1910, the Haar **mother** and **father**[2] wavelets are an orthonormal basis for $\ell^2(\mathbb{Z}_4)$ in the following sense.

DEFINITION 31. We say that $H_M \in \ell^2(\mathbb{Z}_4)$ given by

$$H_M = \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, 0, 0\right)$$

is the **mother Haar wavelet**. We say that $H_F \in \ell^2(\mathbb{Z}_4)$ given by

$$H_F = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 0, 0\right)$$

is the **father Haar wavelet**.

THEOREM 3.1. The set[3]

$$\{R_{2k}H_M\}_{k=0}^{N/2-1} \cup \{R_{2k}H_F\}_{k=0}^{N/2-1}$$

is an orthonormal basis for $\ell^2(\mathbb{Z}_4)$.

PROOF. The proof is elementary: it's clear that $\langle H_M, H_M\rangle = \langle H_F, H_F\rangle = 1$. Furthermore, because of their construction, it's obvious that both H_M and H_F are orthogonal to R_2H_M and R_2H_F. Finally,

$$\langle H_M, H_F\rangle = \frac{\sqrt{2}}{2}\cdot\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}\cdot -\frac{\sqrt{2}}{2} + 0\cdot 0 + 0\cdot 0 = 0.$$

The remainder of the proof is left to the reader. □

[2]We will define these terms in general later on.

[3]This notation looks excessive and obfuscating for such a small set. Later on, we'll use the same notation for arbitrarily large sets of wavelets.

What happens when we project a signal onto the Haar basis? Just as we had hoped, we learn a little bit about the frequency properties of a signal and lose a little bit of information about the spatial properties of a signal.

First, the Haar basis is somewhat spatially localized: if we project the standard basis for $\ell^2(\mathbb{Z}_4)$ onto the Haar basis, then we can spatially differentiate between $\mathbf{e}_0$ and $\mathbf{e}_2$, but not between $\mathbf{e}_0$ and $\mathbf{e}_1$. That is, $\mathbf{e}_0$ and $\mathbf{e}_1$ can be rewritten as sums of *only* the left-hand members of the Haar basis—H_M and H_F. Similarly, $\mathbf{e}_2$ and $\mathbf{e}_3$ can be rewritten as sums of *only* the right-hand members of the Haar basis—R_2H_M and R_2H_F.

Second, the Haar basis is localized in the frequency domain. Consider the Fourier transforms of the basis vectors (we've already computed the Fourier transform of H_M):

$$|\mathcal{F}(H_M)| = \left(0, 1, \sqrt{2}, 1\right)$$
$$|\mathcal{F}(H_F)| = \left(\sqrt{2}, 1, 0, 1\right).$$

In fact, H_M and H_F are high and low-pass filters, respectively[4]. When we project a signal onto these basis vectors, we know from Equation 2.7 that we're really convolving our signal with a high and low-pass filters[5].

Example 17. Consider the signal $v \in \ell^2(\mathbb{Z}_4)$ such that

$$v = (5, 4, 0, 0).$$

Then $\langle R_2H_M, v\rangle = \langle R_2H_F, v\rangle = 0$. However,

[4]This might be hard to see. Keep two things in mind as you interpret them. First, remember that the Fourier transform is symmetric—the highest frequencies are actually low frequencies. Second, note that the Haar wavelets are pretty terrible filters.

[5]Note: this isn't the whole story. Equation 2.7 is for bases of the form $\{R_k w\}_{k=0}^{N-1}$; we have two bases of the form $\{R_{2k} w\}_{k=0}^{N/2-1}$. We'll be more careful about exactly what we're doing here later on. For the moment, it's just important that we make this connection between wavelets, convolution, and bases.

$$\langle H_F, v\rangle = 5\cdot\frac{\sqrt{2}}{2} + 4\cdot\frac{\sqrt{2}}{2} = \frac{9\sqrt{2}}{2}$$
$$\langle H_M, v\rangle = 5\cdot\frac{\sqrt{2}}{2} - 4\cdot\frac{\sqrt{2}}{2} = \frac{\sqrt{2}}{2}$$

If $W = \{R_{2k}H_F\}_{k=0}^{N/2-1} \cup \{R_{2k}H_M\}_{k=0}^{N/2-1}$, then

$$[v]_W = \left(\frac{9\sqrt{2}}{2}, 0, \frac{\sqrt{2}}{2}, 0\right).$$

This is the **first-stage discrete Haar wavelet transform** of v. The Appendix of this thesis contains code for the complete Haar wavelet transform—we will explain what a "complete" wavelet transform means in the upcoming sections.

1.2. Wavelet Transform Algorithm Outline. We now present an outline of the Discrete Wavelet transform in general; this outline will serve as a guide for understanding the aim of the mathematics we will soon present.

Suppose that a discrete signal z supports frequencies[6] from 0 to p.

- First stage decomposition.
  - Use a **high-pass** filter on z; call this new signal z_H.
    - Signal z_H only contains frequencies from $p/2$ to p.
    - Downsample by removing every other term from z_H[7].
  - Use a **low-pass** filter on z; call this new signal z_L.
    - Signal z_L contains only frequencies from 0 to $p/2$.
    - Downsample by removing every other term.
- Second stage decomposition.
  - Use a **high-pass** filter on z_L; call this new signal z_{L_H}.
    - Signal z_{L_H} contains only frequencies from $p/4$ to $p/2$.
    - Downsample by removing every other term.
  - Use a **low-pass** filter on z_L; call this new signal z_{L_L}.

[6]Refer back to the discussion in Chapter 2 on aliasing.
[7]That is, for $v \in \ell^2(\mathbb{Z}_4)$ we would remove the second and fourth terms from the signal.

  * Signal z_{L_L} contains only frequencies from 0 to $p/4$.
  * Downsample by removing every other term.

- Iterate.

Momentarily ignore the theoretical necessity of downsampling; instead, consider the kind of spatial and frequency information we know about the signal at each stage. Before we began filtering our signal z, we had perfect spatial resolution on z, but we knew only that z contained frequencies from 0 to p. After passing z through a high-pass filter and downsampling, we exchanged some spatial resolution for some frequency resolution in our new "half-signal", z_H. Because of the downsampling[8], we lost half of our spatial resolution, but we doubled our frequency resolution—the frequencies in z_H are between $p/2$ and p. The same goes for our lower "half-signal" z_L: we lost half of our spatial resolution, but we know that the frequencies in z_L are between 0 and $p/2$.

At each stage, the traditional wavelet transform only acts on the lower "half-signals". This is a convention, rather than a mathematical necessity; in general, we want more spatial resolution on high frequencies but more frequency resolution on low frequencies. We will stick with this convention for the moment; however, the way that we decompose a signal can be adapted to match our application.

Let me now explicitly compare the discrete Haar transform for $\ell^2(\mathbb{Z}_4)$ with the general first-stage transform outlined above. First, if we directly convolve a signal $v \in \ell^2(\mathbb{Z}_4)$ with the high-pass filter/mother Haar wavelet H_M, then, according to Equation 2.7, we will have the projection of v onto $\{R_k H_M\}_{k=0}^{N-1}$, that is,

$$v * H_M = [v]_{\{R_k H_M\}_{k=0}^{N-1}}.$$

But we don't want the whole projection—since the "mother" half of the basis is the set $\{R_{2k} H_M\}_{k=0}^{N/2-1}$, we only want *every other* term of $[v]_{\{R_k H_M\}_{k=0}^{N-1}}$. This is why downsampling is necessary.

[8]This is sort of misleading—the extra terms are redundant with respect to the wavelet basis, but, as it turns out, they also won't tell us anything else spatially about our signal.

2. Wavelet Bases

Here are the questions you might be asking at this point:

(1) How can we construct different kinds of wavelet bases?
(2) Why do we construct wavelet bases for $\ell^2(\mathbb{Z}_N)$ in the form $\{R_{2k}v\}_{k=0}^{N/2-1} \cup \{R_{2k}u\}_{k=0}^{N/2-1}$ for some high and low pass filters v and u?

That's what this section will be about. It's going to take a lot of proofs and a lot of time—that's why I tried to prime the reader for this section with the discrete Haar wavelet transform and a general outline of discrete wavelet transforms.

Recall the vectors u_1 and u_2 that we introduced in the beginning of the chapter. Each of these vectors had frequency localized properties: we identified them as half of the discrete Haar wavelet basis *and* as high-pass filters. In earlier chapters, we've also noted that the standard basis can be thought of as the basis generated by $\mathbf{e}_0$, that is, $\{R_k\mathbf{e}_0\}_{k=0}^{N-1}$—this basis has perfect spatial resolution. What happens if we try to construct an orthogonal basis of a similar form, but such that each basis vector has frequency localized properties?

Example 18. What does a set of the form I've just described look like for u_1? That is, what does $\{R_k u_1\}_{k=0}^{N-1}$ look like?

$$R_0u_1 = \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, 0, 0\right)$$
$$R_1u_1 = \left(0, \frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, 0\right)$$
$$R_2u_1 = \left(0, 0, \frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)$$
$$R_3u_1 = \left(-\frac{\sqrt{2}}{2}, 0, 0, \frac{\sqrt{2}}{2}\right)$$

The set above has frequency localized properties. However, it isn't an orthogonal basis, and it turns out that no set of this form can simultaneously have frequency localized properties *and* be an orthogonal basis. We now prove this; it is a big result, and it critically constrains the way wavelets can be constructed.

THEOREM 3.2. Let $w \in \ell^2(\mathbb{Z}_N)$. Then $\{R_k w\}_{k=0}^{N-1}$ is an orthonormal basis for $\ell^2(\mathbb{Z}_N)$ if and only if $|\mathcal{F}(w)(n)| = 1$ for all $n \in \mathbb{Z}_N$.

PROOF. Consider the following claim: it turns out that the set $\{R_k w\}_{k=0}$ is an orthonormal basis for $\ell^2(\mathbb{Z}_N)$ if and only if[9]

$$\langle w, R_k w \rangle = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k = 1, 2, ..., N-1. \end{cases}$$

That is, $\langle w, R_k w \rangle = \mathbf{e}_0$. But from 3.5, we know that $w * \tilde{w} = \mathbf{e}_0$. Furthermore, it turns out that $\mathcal{F}(\tilde{w}) = \overline{\mathcal{F}(w)}$.

We proved in Theorem 2.4 that $|\mathcal{F}(\mathbf{e}_0)(m)| = 1$ for $m = 0, 1, ..., N-1$.

Then from the Convolution Theorem (Theorem 2.6) we know that

$$\begin{aligned} \mathcal{F}(w * \tilde{w}) &= \mathcal{F}(\mathbf{e}_0) \\ \mathcal{F}(w) \cdot \mathcal{F}(\tilde{w}) &= 1 \\ \mathcal{F}(w) \cdot \overline{\mathcal{F}(w)} &= 1 \\ |\mathcal{F}(w)|^2 &= 1. \end{aligned}$$

This ends our proof.

□

[9]Proof sketch: combinatorically speaking, there should be a lot of vectors to check when we're proving that a basis is orthonormal (it should take $\binom{N}{2} + N$ inner products). However, it should be easy but tedious to show that $\langle R_k w, R_{k+j} w \rangle = \langle w, R_j w \rangle$ for all j and k. This should prove the claim.

Recall Theorem 2.4; it demonstrated that the standard basis had no frequency resolution because $|\mathcal{F}(\mathbf{e}_i)(n)| = 1$ for each basis vector. Theorem 3.2 is more general but similar in its conclusion: *no* orthogonal basis of the form $\{R_k w\}$ can have frequency localized properties.

However, we can come up with an orthogonal basis with somewhat frequency localized properties *and* somewhat spatially localized properties if we take a slightly different approach. The next couple of theorems demonstrate the plausibility of this method. Before we introduce the main theorems, however, we have to introduce a small definition and prove something about it. The only utility of this theorem is for theorems that follow it; it makes the proofs a little less messy.

Definition 32. For $z \in \ell^2(\mathbb{Z}_N)$, define $z^\times \in \ell^2(\mathbb{Z}_N)$ by

$$z^\times(n) = (-1)^n z(n) \text{ for all } n.$$

Furthermore, note that

$$(z + z^\times)(n) = z(n)(1 + (-1)^n) = \begin{cases} 2z(n) & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd.} \end{cases} \tag{3.1}$$

Theorem 3.3. Suppose $M \in \mathbb{N}, N = 2M$, and $z \in \ell^2(\mathbb{Z}_N)$. Then

$$\mathcal{F}(z^\times)(n) = \mathcal{F}(z)(n + M) \text{ for all } n. \tag{3.2}$$

Proof. By definition,

$$\begin{aligned}
\mathcal{F}(z^{\times})(n) &= \sum_{k=0}^{N-1} z^{\times}(k)e^{-2\pi ikn/N} \\
&= \sum_{k=0}^{N-1} (-1)^k z(k)e^{-2\pi ikn/N} \\
&= \sum_{k=0}^{N-1} z(k)e^{-i\pi k}e^{-2\pi ikn/N} \\
&= \sum_{k=0}^{N-1} z(k)e^{-2\pi ik(n+N/2)/N} \\
&= \sum_{k=0}^{N-1} z(k)e^{-2\pi ik(n+M)/N} \\
&= \mathcal{F}(z)(n+M)
\end{aligned}$$

□

THEOREM 3.4. Suppose $M \in \mathbb{N}$, $N = 2M$, and $w \in \ell^2(\mathbb{Z}_N)$. Then $\{R_{2k}w\}_{k=0}^{M-1}$ is an orthonormal set with M elements if and only if

$$|\mathcal{F}(w)(n)|^2 + |\mathcal{F}(w)(n+M)|^2 = 2 \text{ for } n = 0, 1, ..., M-1.$$

PROOF. We know from Equation 3.1 that

$$(z + z^{\times})(n) = \begin{cases} 2z(n) & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd.} \end{cases}.$$

Since $w * \tilde{w}$ is a member of $\ell^2(\mathbb{Z}_N)$,

$$((w * \tilde{w}) + (w * \tilde{w})^{\times})(n) = \begin{cases} 2(w * \tilde{w})(n) & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd.} \end{cases}$$

Now consider the following claim: the set $\{R_{2k}w\}_{k=0}^{M-1}$ is an orthonormal set if and only if[10]

$$\langle w, R_{2k}w\rangle = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k = 1, 2, ..., M-1. \end{cases}$$

By adapting Theorem 2.8 slightly, it's straightforward to see that $(w * \tilde{w})(2k) = \langle w, R_{2k}w\rangle$. That is, we're only interested in the *even* members of the convolution. Putting everything together, the set $\{R_{2k}w\}_{k=0}^{M-1}$ is an orthonormal set if and only if

$$((w * \tilde{w}) + (w * \tilde{w})^{\times})(2k) = 2(w * \tilde{w})(2k) = 2\left\langle w, R_{2k}w\right\rangle = \begin{cases} 2 & \text{if } k = 0 \\ 0 & \text{if } k = 1, 2, ..., M-1. \end{cases}$$

This is equivalent to saying

$$((w * \tilde{w}) + (w * \tilde{w})^{\times})(2k) = 2\mathbf{e}_0.$$

We now find the Fourier transform of both sides of this equation. Due to the linearity properties of the Fourier transform,

$$\mathcal{F}(2\mathbf{e}_0)(n) = 2\mathcal{F}(\mathbf{e}_0)(n) = 2 \tag{3.3}$$

for all n (we've already proved that $\mathcal{F}(\mathbf{e}_0)(n) = 1$ for all n). By the convolution theorem,

$$\mathcal{F}(w * \tilde{w})(n) = \mathcal{F}(w)(n) \cdot \mathcal{F}(\tilde{w})(n) = \mathcal{F}(w)(n) \cdot \overline{\mathcal{F}(w)(n)} = |\mathcal{F}(w)(n)|^2. \tag{3.4}$$

By Equation 3.2, we know

$$\mathcal{F}((w * \tilde{w})^{\times})(n) = \mathcal{F}(w * \tilde{w})(n + M) = |\mathcal{F}(w)(n + M)|^2. \tag{3.5}$$

[10]Proof outline: use the same sort of techniques that you used earlier to prove a very similar claim.

Putting Equations 3.3, 3.4, and 3.5 together,

$$((w * \tilde{w}) + (w * \tilde{w})^{\times})(2k) = 2\mathbf{e}_0$$

if and only if

$$|\mathcal{F}(w)(n)|^2 + |\mathcal{F}(w)(n+M)|^2 = 2 \text{ for } n = 0, 1, ..., M-1.$$

This concludes our proof. □

Theorem 3.4 indicates necessary conditions for constructing an orthonormal set of vectors in $\ell^2(\mathbb{Z}_N)$—note that since $M = N/2$, this set alone cannot be a basis for $\ell^2(\mathbb{Z}_N)$. Note that the mother and father Haar wavelets satisfy the conditions specified by Theorem 3.4. As the next definition indicates, if we are able to construct *two* of these sets, which are also orthogonal to one another, then the union of these sets will be an orthogonal basis for $\ell^2(\mathbb{Z}_N)$.

Definition 33. Suppose N is an even integer, say $N = 2M$ for some $M \in \mathbb{N}$. An orthonormal basis for $\ell^2(\mathbb{Z}_N)$ of the form

$$\{R_{2k}u\}_{k=0}^{M-1} \cup \{R_{2k}v\}_{k=0}^{M-1}$$

for some $u, v \in \ell^2(\mathbb{Z}_N)$, is called a first-stage wavelet basis for $\ell^2(\mathbb{Z}_N)$. We call u and v the generators of the first-stage wavelet basis. We sometimes also call u the **father wavelet** and v the **mother wavelet**[11].

We now present a few criteria for constructing wavelet bases.

Theorem 3.5. Suppose $M \in \mathbb{N}$ and $N = 2M$. Let $u, v \in \ell^2(\mathbb{Z}_N)$. Then

$$W = \{R_{2k}v\}_{k=0}^{M-1} \cup \{R_{2k}u\}_{k=0}^{M-1}$$

is an orthonormal basis for $\ell^2(\mathbb{Z}_N)$ if and only if

[11]The "mother" denotation makes more sense in the context of the continuous wavelet transform. The idea is that we're projecting a function onto scaled and shifted editions of the same function—the "mother" of other functions. At this point, we've talked about the shifting aspect; we'll get to the scaling part later.

$$|\mathcal{F}(u)(m)|^2 + |\mathcal{F}(u)(m+M)|^2 = 2 \tag{3.6}$$

$$|\mathcal{F}(v)(m)|^2 + |\mathcal{F}(v)(m+M)|^2 = 2 \tag{3.7}$$

and

$$\mathcal{F}(u)(m) \cdot \overline{\mathcal{F}(v)(m)} + \mathcal{F}(u)(m+M) \cdot \overline{\mathcal{F}(v)(m+M)} = 0 \tag{3.8}$$

PROOF. From Theorem 3.4, we know that Equations 3.6 and 3.7 ensure that $\{R_{2k}v\}_{k=0}^{M-1}$ and $\{R_{2k}u\}_{k=0}^{M-1}$ are both orthonormal sets. We need to verify that the condition specified by Equation 3.8 ensures that $\{R_{2k}v\}_{k=0}^{M-1} \cup \{R_{2k}u\}_{k=0}^{M-1}$ is an orthonormal set. We claim that we only need to check that

$$\langle R_{2k}u, R_{2j}v\rangle \text{ for all } j,k = 0,1,...,M-1 \tag{3.9}$$

when Equation 3.8 holds.

We leave it to the reader to show that Equation 3.9 is equivalent to[12]

$$\langle u, R_{2j}v\rangle = 0 \text{ for all } j = 0,1,...,M-1.$$

By Theorem 2.8, we know that

$$\langle u, R_{2j}v\rangle = (u * \tilde{v})(2k).$$

We also know that $(u * \tilde{v})(2k) = 0$ for $k = 0,1,...,M-1$ if and only if

$$(u * \tilde{v})(k) + (u * \tilde{v})^{\times}(k) = 0 \text{ for } k = 0,1,...,N-1$$

since, by Equation 3.1, the expression given above will be always be zero when k is odd.

[12]This is very similar to the other proofs in this section that we've asked the reader to complete. Note that when $k = j$, $\langle R_{2k}u, R_{2j}v\rangle = \langle u, v\rangle$.

Then by Theorems 2.7, 3.3, and the convolution theorem,

$$(u * \tilde{v})(k) + (u * \tilde{v})^{\times}(k) = 0$$

$$\mathcal{F}((u * \tilde{v}))(m) + \mathcal{F}((u * \tilde{v})^{\times})(m) = 0$$

$$\mathcal{F}(u)(m)\overline{\mathcal{F}(v)(m)} + \mathcal{F}(u + M)\overline{\mathcal{F}(v + M)} = 0$$

Since $\{R_{2k}v\}_{k=0}^{M-1} \cup \{R_{2k}u\}_{k=0}^{M-1}$is an orthonormal set of size N, it is a basis for $\ell^2(\mathbb{Z}_N)$.

□

EXAMPLE 19. Suppose that N is divisible by 4. Define $\mathcal{F}(u), \mathcal{F}(v) \in \ell^2(\mathbb{Z}_N)$ by

$$\mathcal{F}(u)(n) = \begin{cases} \sqrt{2} & \text{for } n = 0, 1, ..., \frac{N}{4} - 1 \text{ or } n = \frac{3N}{4}, \frac{3N}{4} + 1, ..., N - 1 \\ 0 & \text{for } n = \frac{N}{4}, \frac{N}{4} + 1, ..., \frac{3N}{4} - 1 \end{cases}$$

and

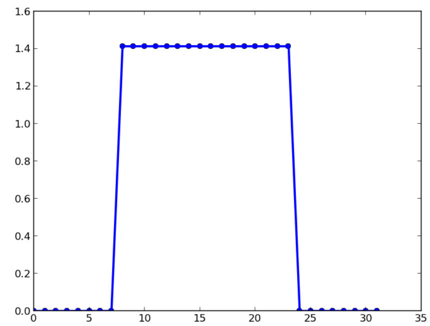
$$\mathcal{F}(v)(n) = \begin{cases} 0 & \text{for } n = 0, 1, ..., \frac{N}{4} - 1 \text{ or } n = \frac{3N}{4}, \frac{3N}{4} + 1, ..., N - 1 \\ \sqrt{2} & \text{for } n = \frac{N}{4}, \frac{N}{4} + 1, ..., \frac{3N}{4} - 1 \end{cases}$$

When we compute the IDFT transform of $\mathcal{F}(u)$ and $\mathcal{F}(v)$, u and v will be generators of a first-stage wavelet basis; we call these the set of **Shannon wavelets**. We plot a representative Shannon mother wavelet in Figure 3.1. It is left to the reader to check that the construction above satisfies the criteria given by Theorem 3.5. In addition, note that u and v will be high and low-pass filters, respectively[13].

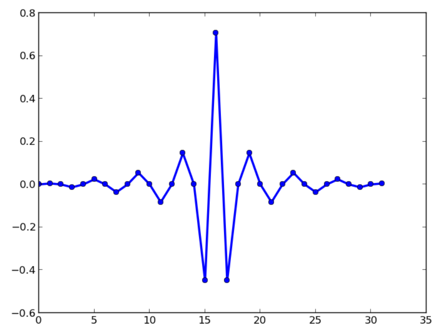
THEOREM 3.6. Suppose $M \in \mathbb{N}, N = 2M$, and $u \in \ell^2(\mathbb{Z}_N)$ is such that $\{R_{2k}u\}_{k=0}^{M-1}$ is an orthonormal set with M elements. Define $v \in \ell^2(\mathbb{Z}_N)$ by

$$v(m) = (-1)^m\overline{u(1 - m)}$$

[13] Again, the slightly non-intuitive construction is the result of aliasing.



(a) Frequncy-Amplitude Plot



(b) Time-Amplitude Plot

FIGURE 3.1. Shannon Mother Wavelet with $N = 32$

for all m. Then $\{R_{2k}u\}_{k=0}^{M-1} \cup \{R_{2k}v\}_{k=0}^{M-1}$ is a first-stage wavelet basis for $\ell^2(\mathbb{Z}_N)$.

We mentioned earlier that we would need to downsample in order to compute the wavelet transform; we now define downsampling mathematically.

DEFINITION 34. Let $M \in \mathbb{N}$; let $N = 2M$. Then we define the **downsampling** operator $D : \ell^2(\mathbb{Z}_N) \to \ell^2(\mathbb{Z}_M)$ by

$$D[z](n) = z(2n) \text{ for } n = 0, 1, ..., M-1.$$

EXAMPLE 20. Let $z = [1, 2, 3, 4]$. Then $D[z] = [1, 3]$.

Definition 35. Let $M \in \mathbb{N}$; let $N = 2M$. Then we define the **upsampling** operator $U : \ell^2(\mathbb{Z}_M) \to \ell^2(\mathbb{Z}_N)$ by

$$U[z](n) = \begin{cases} z(n/2) & \text{if } n \text{ is even,} \\ 0 & \text{if } n \text{ is odd.} \end{cases}$$

At this point, we're ready to actually compute a wavelet transform of a signal.

Definition 36. Suppose that for $u, v \in \ell^2(\mathbb{Z}_N)$, $\{R_{2k}u\}_{k=0}^{M-1} \cup \{R_{2k}v\}_{k=0}^{M-1}$ is an orthonormal basis for $\ell^2(\mathbb{Z}_N)$. Then we define the **first-stage discrete wavelet transform** of $z \in \ell^2(\mathbb{Z}_N)$ by

$$\mathcal{W}[z] = \{x_1, y_1\}$$

where

$$x_1 = D[z * \tilde{u}]$$

and

$$y_1 = U[z * \tilde{v}].$$

It should be clear from Theorem 2.8 that together, x_1 and y_1 constitute the projection of z onto the first-stage wavelet basis generated by u and v. That is, $x_1, y_1 \in \ell^2(\mathbb{Z}_M)$, $x_1(i) = \langle z, R_{2i}u \rangle$ and $y_1(i) = \langle z, R_{2i}v \rangle$. Often we concatenate y_1 with x_1; the resultant vector is an element of $\ell^2(\mathbb{Z}_N)$. However, it's easier to manipulate x_1 and y_1 separately.

When we introduced the DFT, we also introduced the IDFT - we now provide the analogous inversion formula for the wavelet transform.

Example 21. Let $z = [1, 3]$. Then $U[z] = [1, 0, 3, 0]$.

Theorem 3.7. Let $M \in \mathbb{N}$, let $N = 2M$, and let $u, v \in \ell^2(\mathbb{Z}_N)$ such that $\{R_{2k}u\}_{k=0}^{M-1} \cup \{R_{2k}v\}_{k=0}^{M-1}$ is an orthonormal basis for $\ell^2(\mathbb{Z}_N)$. Then

$$u * U[D[z * \tilde{u}]] + v * U[D[z * \tilde{v}]] = z.$$

That is,

$$u * U[x_1] + v * U[y_1] = z.$$

PROOF. Let $z \in \ell^2(\mathbb{Z}_N)$. If $\{R_{2k}u\}_{k=0}^{M-1} \cup \{R_{2k}v\}_{k=0}^{M-1}$ is an orthonormal basis for $\ell^2(\mathbb{Z}_N)$ then we know that

$$z = \sum_{i=0}^{M-1} \langle z, R_{2i}u\rangle R_{2i}u + \sum_{j=0}^{M-1} \langle z, R_{2j}v\rangle R_{2j}v.$$

We now show that $u * U[D[z * \tilde{u}]] = \sum_{i=0}^{M-1} \langle z, R_{2i}u\rangle u$. Remember that $D[z * \tilde{u}](i) = \langle z, R_{2i}u\rangle$. Then

$$(u * U[D[z * \tilde{u}]])(k) = \sum_{i=0}^{N-1} U[D[z * \tilde{u}]](i) \cdot u(k - i).$$

Note that for odd k, $U[D[z * \tilde{u}]](k) = 0$. Therefore,

$$\begin{aligned}\sum_{i=0}^{N-1} U[D[z * \tilde{u}]](i) \cdot u(k - i) &= \sum_{i=0}^{M-1} \langle z, R_{2i}u\rangle \cdot u(k - 2i) \\ &= \sum_{i=0}^{M-1} \langle z, R_{2i}u\rangle \cdot R_{2i}u(k).\end{aligned}$$

□

2.1. Wavelet Transform Iteration.

DEFINITION 37. Suppose N is divisible by 2^p for $p \in \mathbb{N}$. We define a pth stage wavelet filter sequence as a sequence of vectors $u_1, v_1, u_2, v_2, ..., u_p, v_p$ such that, for each $\ell \in \{1, 2, ..., p\}$,

$$u_\ell, v_\ell \in \ell^2(\mathbb{Z}_{N/2^{\ell-1}}).$$

EXAMPLE 22. Though we're confident the reader followed the previous definition, keeping track of the lengths of the filters while trying to understand the upcoming definitions can be a challenge. Let $N = 8$. Then

$$u_1, v_1 \in \ell^2(\mathbb{Z}_8)$$

$$u_2, v_2 \in \ell^2(\mathbb{Z}_4)$$

$$u_3, v_3 \in \ell^2(\mathbb{Z}_2).$$

DEFINITION 38. Suppose N is divisible by 2^p for $p \in \mathbb{N}$. Then for $z \in \ell^2(\mathbb{Z}_N)$, define

$$x_1 = D(z * \tilde{v}_1) \in \ell^2(\mathbb{Z}_{N/2})$$

and

$$y_1 = D(z * \tilde{u}_1) \in \ell^2(\mathbb{Z}_{N/2}).$$

Furthermore, we recursively define $x_2, y_2, ..., x_p, y_p$ by

$$x_\ell = D(y_{\ell-1} * \tilde{v}_\ell) \in \ell^2(\mathbb{Z}_{N/2^\ell})$$

and

$$y_\ell = D(y_{\ell-1} * \tilde{v}_\ell) \in \ell^2(\mathbb{Z}_{N/2^\ell})$$

The pth stage wavelet transform of z is the set of vectors $\{y_p, x_p, ..., x_2, x_1\}$.

EXAMPLE 23. As an extension of the example provided earlier in this chapter, consider the second stage Haar wavelet transform of $v = (5, 4, 0, 0)$. We already know that $x_1 = \left(\frac{9\sqrt{2}}{2}, 0\right)$ and $y_1 = \left(\frac{\sqrt{2}}{2}, 0\right)$. Then

$$x_2 = \langle H_M, y_1 \rangle = \frac{9\sqrt{2}}{2} \cdot \frac{\sqrt{2}}{2} + 0 \cdot \frac{\sqrt{2}}{2}$$
$$y_2 = \langle H_F, y_1 \rangle = \frac{9\sqrt{2}}{2} \cdot \frac{\sqrt{2}}{2} - 0 \cdot \frac{\sqrt{2}}{2}.$$

Then the second stage Haar wavelet transform of v is the set $\{y_2, x_2, x_1\}$.

3. Significance Tests Using the Haar Wavelet Transform

We now introduce the first of four significance tests in this thesis. Though different tests should be used in different contexts, all of the significance tests aim to indicate when a given signal is *not* produced by white noise. Why is this important, and what does this mean?

Suppose we are examining some signal - a string of data $z(t)$ formed by taking discrete samples at even time intervals of some process.

For example, suppose $z(t)$ represents the temperature of a lake at a given time t. We might suspect that there is a pattern to our data— we should expect periodic behavior at daily and yearly time scales. Thus, we can say that

$$z(t) = f(t) + W(t),$$

where $f(t)$ is the true temperature of the lake at time t and $W(t)$ is measurement error at time t[14]. When we don't *a priori* know the form of $f(t)$[15], then we will have to establish mathematical tests to differentiate $f(t)$ from $W(t)$.

In our case, we're interested in ecological signals, which, for many reasons, we suspect have transient behavior. As such, we analyze them using wavelets. Therefore, we want to establish tests that can tell us, with confidence, which wavelet transform coefficients of our signal $z(t)$ can be attributed to a white noise signal, $W(t)$.

The following two tests are my own, although they are modeled after the tests found in the next chapter.

3.1. One Signal Significance with the Haar Wavelet Transform. Suppose we want to determine if a signal $z \in \ell^2(\mathbb{Z}_N)$ is white noise by examining the coefficients of the wavelet transform of z. We first find the Haar wavelet transform of z; we'll treat the set of corresponding transform coefficients $\{x_1, x_2, ..., x_p, y_p\}$ as a long vector $H[z] \in \ell^2(\mathbb{Z}_N)$.

[14]In general, we will assume that $W(t)$ is white noise.

[15]In practice, of course, we won't ever know the form of $f(t)$. In fact, we don't even know that $f(t)$ is non-zero, or that $f(t)$ is periodic or stationary. This, perhaps, is the fundamental strength of the wavelet transform: we don't have to make as many assumptions about the underlying function $f(t)$.

We derive the probability distribution for each discrete Haar wavelet transform coefficient of a white noise signal; afterwards, we'll demonstrate how we can use this information to conduct a significance test.

THEOREM 3.8. Let $N = 2^n$ for some $n \in \mathbb{N}$. For a white noise signal with variance σ^2, $W \in \ell^2(\mathbb{Z}_N)$, let $H[W] \in \ell^2(\mathbb{Z}_N)$ be the discrete Haar wavelet transform of W. Then $H[W](i)$ has a normal distribution $N(0, \sigma^2)$.

PROOF. We prove the theorem inductively; we follow the algorithm given by the Python program given in the Appendix.

We denote the j^{th} stage discrete Haar wavelet transform by $H_j[W]$. Consider the members of the first stage Haar wavelet transform; that is,

$$H_1[W](i) = \begin{cases} \frac{W(i)+W(i+1)}{\sqrt{2}} & \text{for } i = 0, 1, ..., \frac{N}{2} - 1 \\ \frac{W(i)-W(i+1)}{\sqrt{2}} & \text{for } i = \frac{N}{2}, ..., N - 1. \end{cases}$$

From Theorem 1.10, for $i = 0, 1, ..., \frac{N}{2} - 1$,

$$\begin{aligned} \mathbf{Var}[H_1[W](i)] &= \mathbf{Var}\left[\frac{W(i) + W(i+1)}{\sqrt{2}}\right] \\ &= \frac{1}{2}\mathbf{Var}[W(i) + W(i+1)] \\ &= \sigma^2 \end{aligned}$$

and for $i = \frac{N}{2}, ..., N - 1$,

$$\begin{aligned} \mathbf{Var}[H_1[W](i)] &= \mathbf{Var}\left[\frac{W(i) - W(i+1)}{\sqrt{2}}\right] \\ &= \frac{1}{2}\mathbf{Var}[W(i) - W(i+1)] \\ &= \sigma^2. \end{aligned}$$

Assume that for each i, $H_j[W](i)$ is a normally distributed random variable with mean $\mu = 0$ and variance σ^2. Then

$$H_{j+1}[W](i) = \begin{cases} \frac{H_j[W](i)+H_j[W](i+1)}{\sqrt{2}} & \text{for } i = 0, 1, ..., \frac{N}{2^j} - 1 \\ \frac{H_j[W](i)+H_j[W](i+1)}{\sqrt{2}} & \text{for } i = \frac{N}{2^j}, ..., \frac{N}{2^{j-1}} - 1 \\ H_j[W](i) & \text{for } i = \frac{N}{2^{j-1}}, ..., N-1. \end{cases}$$

Since we assumed that $H_j[W](i)$ is a normally distributed random variable with mean $\mu = 0$ and variance σ^2, then we know that for $i = \frac{N}{2^{j-1}}, ..., N-1$, $H_{j+1}[W](i)$ is a normally distributed random variable with mean $\mu = 0$ and variance σ^2. Then for $i = 0, 1, ..., \frac{N}{2^j} - 1$,

$$\begin{aligned} \mathbf{Var}[H_{j+1}[W](i)] &= \mathbf{Var}\left[\frac{H_j[W](i) + H_j[W](i+1)}{\sqrt{2}}\right] \\ &= \frac{1}{2}\left[H_j[W](i) + H_j[W](i+1)\right] \\ &= \sigma^2. \end{aligned}$$

The distribution of $H_{j+1}[W](i)$ for $i = \frac{N}{2^j}, ..., \frac{N}{2^{j-1}} - 1$ can be proved similarly. □

Theorem 3.9. Let $N = 2^n$ for some $n \in \mathbb{N}$. For a white noise signal $W \in \ell^2(\mathbb{Z}_N)$ with variance $\sigma^2 = 1$, let $H[W] \in \ell^2(\mathbb{Z}_N)$ be the discrete Haar wavelet transform of W. Then

$$H[W](i)^2 \sim \chi_2^2;$$

that is, $H[W](i)^2$ has a chi-square distribution with one degree of freedom.

Proof. This fact follows straightforwardly from the previous proof and the fact that the square of a random variable with a standard normal distribution is chi-square distributed with one degree of freedom. □

We now introduce our significance test.

Procedure 2. **One Signal Haar Wavelet Significance Test**. Suppose $N = 2^p$ for some $p \in \mathbb{N}$ and $z \in \ell^2(\mathbb{Z}_N)$[16]. We want to conduct a test to determine whether the pth stage Haar wavelet transform coefficients of z are significantly different from the pth stage Haar wavelet transform coefficients of a white noise signal $W \in \ell^2(\mathbb{Z}_N)$. We allow for some pre-specified α, the likelihood of a Type I error.

The null hypothesis of our test is that $H[z'](i) \sim \chi_1^2$, where z' is a normalized version of z that we will introduce momentarily; the alternative hypothesis is that $H[z'](i) \not\sim \chi_1^2$. We reject or fail to reject the null hypothesis using the following steps.

(1) Normalize[17] z using Procedure 1 from Chapter 1; we call this normalized signal z'.
(2) Compute the pth stage Haar wavelet transform, $H[z']$, of z'.
(3) Compute the rejection region for the null hypothesis. That is, determine k such that $P(H[z'](i)^2 \geqslant k) = \alpha$, or, equivalently, solve the following equation for k:

$$1 - \int_0^k \frac{1}{2^{1/2}\Gamma(1/2)} x^{\frac{-1}{2}} e^{-\frac{x}{2}} dx = \alpha.$$

(4) For $i = 0, 1, ..., N-1$, if $H[z'](i)^2 \geqslant k$, reject the null hypothesis. Otherwise, we fail to reject the null hypothesis.

We provide examples of the Haar wavelet one-signal significance test in Chapter 5; we also discuss potential problems with its use. We provide the code to compute this test in the Appendix.

When z has some length $N \neq 2$, we make z as periodic as we need it to be, which means that we repeat the entries of z until our signal has length $N' = 2^p$. This adjustment technically renders the significance test invalid, but we suspect that it will nevertheless be ‘accurate.

3.2. Cross Spectrum Significance with the Haar Wavelet Transform. In general, a wavelet cross spectrum aims to determine when two signals have similar wavelet transforms.

[16]We will discuss later what we should do when $N \neq 2^p$.
[17]We do this because we can't make any assumptions about the white noise we're dealing with.

DEFINITION 39. Let $z_1, z_2 \in \ell^2(\mathbb{Z}_N)$. Then we define the **Haar wavelet transform cross spectrum** for $i = 0, 1, ..., N-1$ as

$$H[z_1](i)^2 \cdot H[z_2](i)^2.$$

Suppose both z_1 and z_2 are white noise signals; if we normalize z_1 and z_2, then we know from Theorem 3.9 that for all i, both $H[z_1'](i)$ and $H[z_2'](i)$ will be chi-square distributed with one degree of freedom. To construct a cross spectrum significance test, we want to determine the probability distribution of $H[z_1'](i)^2 H[z_2'](i)^2$.

DEFINITION 40. We say that a random variable X has a Gamma distribution if it has a probability density function of the form

$$f(x; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}.$$

We denote a random variable having a Gamma distribution as $X \sim G(\alpha, \beta)$. Note that if $X \sim G(\nu/2, 2)$, then X has a chi-square distribution with ν degrees of freedom.

THEOREM 3.10. Let $X \sim \chi_1^2$ and $Y \sim \chi_1^2$ be independent. Let $Z = XY$. Then Z has density function

$$f(z) = \frac{1}{\pi\sqrt{z}} K_0(z^{1/2})$$

where K_0 is the modified Bessel function of the second kind of order zero.

PROOF. From [**10**], we cite the probability density function of the product of two independent Gamma distributions. Let $Z \sim G(\alpha_1, \beta_1 = \frac{1}{\alpha_1})$ and $Y \sim G(\alpha_2, \beta_2)$; let X and Y be independent. Then $Z = XY$ is a random variable with probability density function

$$f(z; \alpha_1, \alpha_2, \beta_2) = \frac{2}{z} \left(\frac{\alpha_1 z}{\beta_2}\right)^{\frac{\alpha_1+\alpha_2}{2}} \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} K_{\alpha_2-\alpha_1}\left(2\sqrt{\frac{\alpha_1 z}{\beta_2}}\right),$$

where $K_{\alpha_2-\alpha_1}$ is the modified Bessel function of the second kind of order $\alpha_2-\alpha_1$. If $\alpha_1 = 1/2$, $\alpha_2 = 1/2$, and $\beta_2 = 2$, then X and Y are both chi-square distributed with one degree of freedom. With the appropriate parameter substitutions, it's clear that $Z = XY$ has density function

$$f(z) = \frac{1}{\pi\sqrt{z}} K_0(z^{1/2}).$$

using the fact that $\Gamma\left(\frac{1}{2}\right)^2 = \pi$. □

PROCEDURE 3. **Cross Spectrum Haar Wavelet Significance Test**. Suppose $N = 2^p$ for some $p \in \mathbb{N}$ and $z_1, z_2 \in \ell^2(\mathbb{Z}_N)$. We want to conduct a test to determine whether the cross spectrum coefficients of z_1 and z_2 are significantly different from the pth stage Haar wavelet transform coefficients of two white noise signals $W_1, W_2 \in \ell^2(\mathbb{Z}_N)$. We allow for some pre-specified α, the likelihood of a Type I error.

The null hypothesis of our test is that $H[z'](i) \sim \chi_1^2$, where z' is a normalized version of z that we will introduce momentarily; the alternative hypothesis is that $H[z'](i) \nsim \chi_1^2$. We reject or fail to reject the null hypothesis using the following steps.

(1) Normalize z_1 and z_2 using Procedure 1 from Chapter 1; we call these normalized signals z_1' and z_2'.

(2) Compute the pth stage Haar wavelet transforms, $H[z_1']$ and $H[z_2']$, of z_1' and z_2'.

(3) Compute the rejection region for the null hypothesis. That is, determine k such that $P(H[z_1'](i)^2 H[z_2'](i)^2 \geqslant k) = \alpha$, or, equivalently, solve the following equation for k:

$$1 - \int_0^k \frac{1}{\pi\sqrt{z}} K_0(z^{1/2}) dx = \alpha.$$

(4) For $i = 0, 1, ..., N-1$, if $H[z_1'](i)^2 H[z_2'](i)^2 \geqslant k$, reject the null hypothesis. Otherwise, we fail to reject the null hypothesis.

We provide examples of the Haar wavelet cross spectrum significance test in Chapter 5; we also discuss potential problems with its use. We provide the code to compute this test in the Appendix.

CHAPTER 4

The Continuous Wavelet Transform

We will begin with a brief introduction to continuous wavelet theory, as presented in [**3**] and [**2**]. This section will be remarkably non-technical - we will only need these definitions to define admissibility criteria for wavelets and to get a loose idea about what a wavelet transform is really doing.

To begin our discussion, we first define the set of square-summable functions.

Definition 41.

$$L^2(\mathbb{R}) = \left\{ f : \mathbb{C} \to \mathbb{R} : \int_{-\infty}^{\infty} |f(x)|^2 dx < \infty \right\}$$

We call this the set of **square-summable** functions.

First, note that many common functions are not square-summable; for example,

(1) $f(x) = x$
(2) $g(x) = \sin(x)$
(3) $h(x) = \cos(x)$.

The fact that these last two trigonometric functions are not square-summable turns out to be of fundamental importance to wavelet theory. Fourier analysis relies on these two functions to act as basis vectors for $L^2(0, 2\pi)$ - the set of square-summable functions on the interval $(0, 2\pi)$. But since sine and cosine functions aren't even members of $L^2(\mathbb{R})$, they certainly can't act as basis vectors. Continuous wavelets, however, are square-summable, and just as their discrete counterparts acted as basis vectors for $\ell^2(\mathbb{Z}_N)$, continuous wavelets act as basis vectors for $L^2(\mathbb{R})$.

We require that a wavelet be a *window function*.

DEFINITION 42. A nontrivial function $w \in L^2(\mathbb{R})$ is called a window function if $xw(x)$ is also in $L^2(\mathbb{R})$. The center t^* and radius Δ_w of a window function w are defined to be

$$t^* = \frac{1}{||w||^2} \int_{-\infty}^{\infty} x|w(x)|^2 dx$$

and

$$\Delta_w = \frac{1}{||w||} \sqrt{\int_{-\infty}^{\infty} (x - t^*)^2 |w(x)|^2 dx}$$

In addition, we require that

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\mathcal{F}[\psi](\omega)|^2}{|\omega|} d\omega < \infty,$$

where $\mathcal{F}[\psi]$ is the continuous Fourier transform of ψ.

It turns out that this will be the criterion necessary for a function to be reconstructed from its wavelet transform.

We now introduce the continuous wavelet transform.

DEFINITION 43. Let $f \in L^2(\mathbb{R})$, let ψ be a wavelet. Then we define the continuous wavelet transform of f for $b \in \mathbb{R}$ and $a > 0$ as

$$(W_\psi f)(b, a) = |a|^{1/2} \int_{-\infty}^{\infty} f(x) \overline{\psi\left(\frac{x-b}{a}\right)} dx,$$

We often refer to a as **scale** and b as **shift**.

By the way that we've defined a wavelet, both the wavelet and its Fourier transform will be window functions.

Suppose that a given wavelet ψ and its Fourier transform $\mathcal{F}[\psi]$ have centers and radii given by $t^*, \omega^*, \Delta_\psi$, and $\Delta_{\mathcal{F}[\psi]}$. Then the wavelet transform of a function f evaluated at a and b localizes f within the time-domain within the interval

$$[b + at^* - a\Delta_\psi, b + at^* + a\Delta_\psi].$$

Furthermore, the wavelet transform localizes a function f within the frequency-domain within the interval

$$\left[\frac{\omega^*}{a} - \frac{1}{a}\Delta_{\mathcal{F}[\psi]}, \frac{\omega^*}{a} + \frac{1}{a}\Delta_{\mathcal{F}[\psi]}\right].$$

1. Continuous Wavelet Significance Tests

We now introduce the wavelet transform that we will be using for the remainder of this chapter.

Definition 44. Let $z(t) \in \ell^2(\mathbb{Z}_N)$. Since $z(t)$ is interpreted as a signal sampled at some constant sampling rate, let δt be the time difference between each pair of data points. Let ψ be a wavelet. We define the **sampled continuous wavelet transform** of z for $n = 0, 1, ..., N-1$ and $a > 0$ as

$$T_{n,a}[z] = \sum_{t=0}^{N-1} z(t)\frac{\delta t}{\sqrt{a}}\overline{\psi\left(\frac{(t-n)\delta t}{a}\right)} \tag{4.1}$$

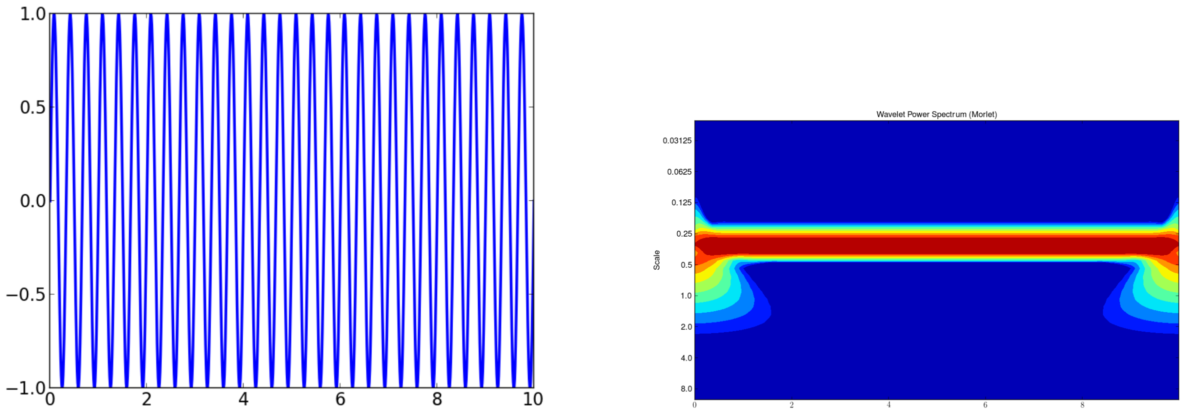
For the remainder of this chapter, we'll just refer to Equation 4.1 as the wavelet transform of $z(t)$ at n and a. Equation 4.1 is a subtle variation of the transform given by Torrence and Compo in [**13**]; reasons for avoiding Torrence and Compo's transform are given in [**6**].

Definition 45. Following [**13**] and [**6**], we define the **wavelet power** of a signal $z(t) \in \ell^2(\mathbb{Z}_N)$ at scale a and shift n as

$$|T_{n,a}[z]|^2.$$

We often plot wavelet power over a domain of n and a; a plot of this form is called a **scaleogram**.

Consider Figure 4.1. On the right, we see a traditional time-amplitude plot of a 3Hz sine function; that is, a plot of $f(t) = \sin(6\pi t)$. On the right, we see the wavelet power spectrum of $f(x)$ plotted in a scaleogram. The scaleogram was constructed using Equation 4.1 and a *Morlet* wavelet; we will introduce the Morlet wavelet momentarily.



(a) Time-Amplitude Plot of $f(t)$

(b) Power Spectrum of $f(t)$

FIGURE 4.1. Plots of $f(t) = \sin(6\pi t)$.

Note that the coefficients are plotted on a spectrum from blue to red; the intensity is proportional to the magnitude of the coefficients. The horizontal axis plots the shift n of the wavelet transform - these coefficients should roughly be interpreted as coefficients in the time domain. The vertical axis corresponds to the wavelet scale parameter a. It turns out that when we use the Morlet wavelet, we can directly compare wavelet scale with Fourier frequency. Finally, note the strange diffusion effects appearing on the extreme left and right boundaries of the transform; these are known as **edge effects**. They are the inevitable consequence of taking the wavelet transform of a finite-length signal; there are some established methods of reducing them.

DEFINITION 46. We say that $\psi(t)$ given below is the **Morlet wavelet**.

$$\psi(t) = \pi^{-1/4} e^{i\omega_0 t} e^{-\frac{t^2}{2}} \tag{4.2}$$

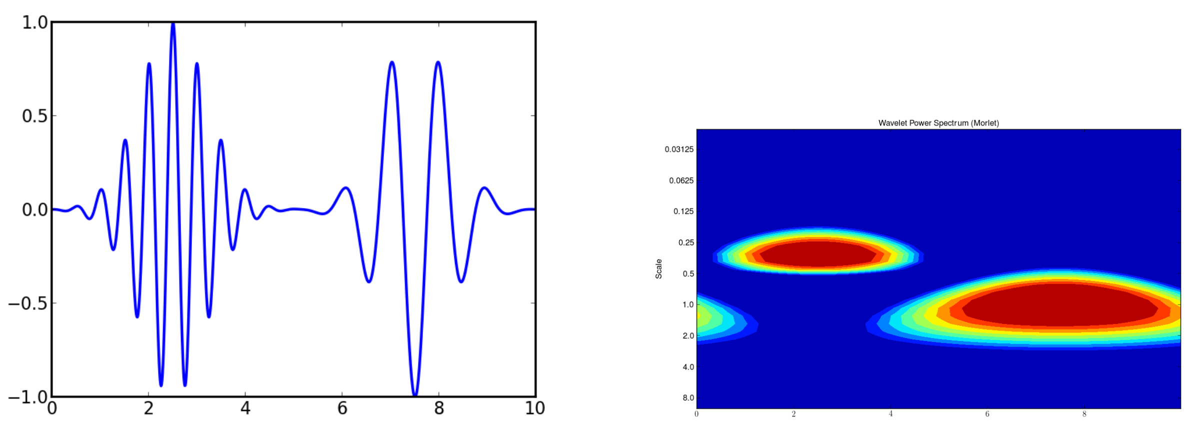
For the remainder of this chapter, $\psi(t)$ will refer to the Morlet wavelet rather than an arbitrary wavelet.

The Morlet wavelet has many desirable properties. In particular, [**13**], [**8**] and [**2**] note that when $\omega_0 = 6.0$, wavelet scale and Fourier frequency are almost exactly inversely proportional[1]; that is, $f = \frac{1}{a}$, where f is Fourier frequency and a is wavelet scale. This greatly eases the interpretation of wavelet transform coefficients, especially - as is the case when the data sets are ecological - when knowledge of the exact frequency being examined is critical to the analysis of results. In addition, the Morlet wavelet has both a real and an imaginary part; a wavelet transform using a Morlet wavelet can be used for phase analysis. Finally, the Morlet wavelet has properties that are useful when analytically deriving the statistical significance tests that we will begin discussing momentarily; see [**6**] and [**7**].

Consider Figure 4.2, a plot of

$$f(t) = e^{-(t-2.5)^2}\cos(2\pi t) + e^{-(t-7.5)^2}\cos(4\pi t)$$

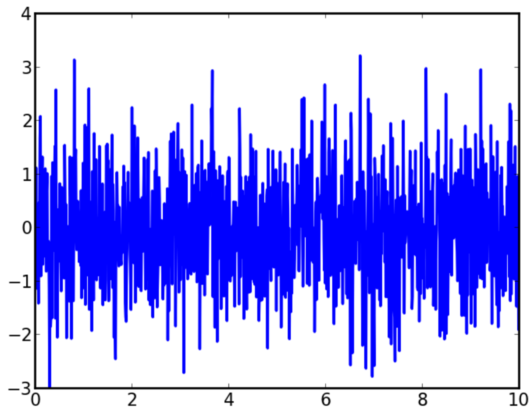
and its wavelet transform. The figure really re-emphasizes the central theme of this thesis; it provides insight into the power of the wavelet transform, and the properties of the wavelet transform that really differentiate it from the Fourier transform - the wavelet transform is able to roughly locate the moment at which the frequency in the sample signal changed.



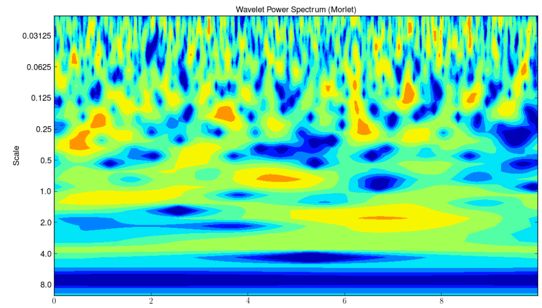
(a) Time-Amplitude Plot of $f(t)$ (b) Power Spectrum of $f(t)$

FIGURE 4.2. Plots of $f(t) = e^{-(t-2.5)^2}\cos(4\pi t) + e^{-(t-7.5)^2}\cos(2\pi t)$.

[1]Looking back at Figure. 4.1, note that the scaleogram has large coefficients when scale $a \approx .33$. This means the peak Fourier frequency is about 3Hz.



(a) Time-Amplitude Plot of $f(t)$



(b) Power Spectrum of $f(t)$

FIGURE 4.3. Plots of $f(t) = W(t)$.

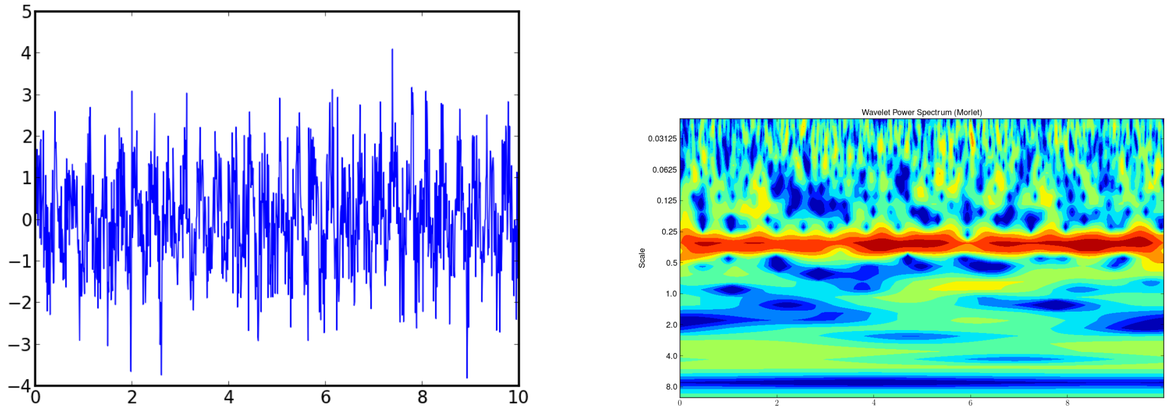
Perhaps the most popular methodological source for the application of the wavelet transform to data sets is given by Torrence and Compo in [**13**]; since its publication, the paper has been cited nearly 5000 times. Though the paper is valuable for its attempt to organize and synthesize years of wavelet research into a relatively straightforward manual for scientists, the work is most notable for its attempt to bring measures of statistical significance[2] to wavelet analysis.

As we noted in Chapter 3, when we compute the wavelet transform of white noise, we see apparently non-random structures within the transform coefficients; see Figure 4.3. Tests for statistical significance are necessary, therefore, to differentiate between features produced by randomness and features truly indicative of some underlying process. Nevertheless, plots like Figure 4.4 allow us to be optimistic about the ability of the wavelet transform to detect some underlying signal $f(t)$; it's actually more obvious that the signal contains a sine function when we look at the wavelet transform of the signal.

Torrence and Compo established tests for statistical significance using Monte Carlo methods; later papers analytically derived the probability distributions necessary for establishing significance tests; see [**6**], [**7**]. Though the findings of [**6**] and [**7**] numerically match the results of Torrence and Compo, we found that the derivations are *wrong*. We will recreate

[2]We will define the general form of statistical significance test in the next section.

their proofs in the following section; we believe that the probability distributions they derive are accurate when δt, the sampling interval, approaches zero.



(a) Time-Amplitude Plot of $f(t)$

(b) Power Spectrum of $f(t)$

FIGURE 4.4. Plots of $f(t) = \sin(6\pi t) + W(t)$.

1.1. One Signal Wavelet Significance Test. Following—and correcting—the derivations found in [**6**], we derive the probability distribution for each Morlet wavelet transform coefficient of a white noise signal; afterwards, we'll demonstrate how we can use this information to conduct a significance test. We begin by proving a series of lemmas.

LEMMA 4.1. Let $W(t) \in \ell^2(\mathbb{Z}_N)$ be a white noise signal with mean $\mu = 0$ and variance σ^2. Then $\Re[T_{n,a}[W]]$ and $\Im[T_{n,a}[W]]$ are normally distributed with mean $\mu = 0$.

PROOF. First, we identify the real and imaginary parts of the wavelet transform:

$$\Re[T_{n,a}[W]] = \sum_{t=0}^{N-1} W(t) \frac{\delta t}{\sqrt{a}} \Re\left[\overline{\psi\left(\frac{(t-n)\delta t}{a}\right)}\right]$$

and

$$\Im[T_{n,a}[W]] = \sum_{t=0}^{N-1} W(t) \frac{\delta t}{\sqrt{a}} \Im\left[\overline{\psi\left(\frac{(t-n)\delta t}{a}\right)}\right].$$

First, note that linear combinations of normally distributed random variables are also normally distributed - both the real and imaginary parts of $T_{n,a}[W]$ given above are linear combinations of normally distributed random variables.

Next, we demonstrate that $\Re[T_{n,a}[W]]$ and $\Im[T_{n,a}[W]]$ have mean $\mu = 0$; that is,

$$\mathbf{E}[\Re[T_{n,a}[W]]] = \mathbf{E}[\Im[T_{n,a}[W]]] = 0.$$

We will only prove the real part; the proof of the imaginary part is identical. By definition,

$$\begin{aligned}
\mathbf{E}[\Re[T_{n,a}[W]]] &= \mathbf{E}\left[\sum_{t=0}^{N-1} W(t)\frac{\delta t}{\sqrt{a}}\Re\left[\overline{\psi\left(\frac{(t-n)\delta t}{a}\right)}\right]\right] \\
&= \frac{\delta t}{\sqrt{a}}\sum_{t=0}^{N-1}\Re\left[\overline{\psi\left(\frac{(t-n)\delta t}{a}\right)}\right]\mathbf{E}\left[W(t)\right] \\
&= 0
\end{aligned}$$

since $\mathbf{E}[W(t)] = 0$ for all t.

□

The following theorem is taken from [**9**] and is necessary for a later proof. Given certain assumptions about X and Y, it provides the criteria necessary to demonstrate that X and Y are independent.

LEMMA 4.2. Let $X_1, X_2, ..., X_n$ be independent normally distributed random variables, and assume that the n^{th} moment of each X_i exists. Let

$$Y_1 = \sum_{i=1}^{n} a_i X_i$$

and

$$Y_2 = \sum_{i=1}^{n} b_i X_i.$$

If

$$\sum_{i=0}^{n} a_i b_i \sigma_i^2 = 0,$$

then X and Y are independent.

The following lemma is *not* true. Nevertheless, we include it: based upon the Monte Carlo results found in [**13**], the overall theorem we will prove using the lemma appears to be approximately accurate for small δt.

LEMMA 4.3. Let $W(t) \in \ell^2(\mathbb{Z}_N)$ be a white noise signal with mean $\mu = 0$ and variance σ^2. Then $\Re[T_{n,a}[W]]$ and $\Im[T_{n,a}[W]]$ are independent.

PROOF. We know $W(t)$'s are mutually independent and normally distributed. Following Lemma 4.2, we want to show that

$$\frac{\delta t^2}{a} \sum_{t=0}^{N-1} \Re\left[\overline{\psi\left(\frac{(t-n)\delta t}{a}\right)}\right] \Im\left[\overline{\psi\left(\frac{(t-n)\delta t}{a}\right)}\right] \mathbf{Var}[W(t)] = 0. \tag{4.3}$$

Since $\mathbf{Var}[W(t)]$ and δt are non-zero, it's apparent that Equation 4.3 only holds when

$$\sum_{t=0}^{N-1} \Re\left[\overline{\left(\frac{(t-n)\delta t}{a}\right)}\right] \Im\left[\overline{\psi\left(\frac{(t-n)\delta t}{a}\right)}\right] = 0.$$

Unfortunately, this just isn't true using the Morlet wavelet. We've written up some simple Python code to calculate the value of the left side of Equation 4.3 in the Appendix, but we'll also simplify the expression to make it more clear that it isn't equal to zero.

Remember that we are using a Morlet wavelet. If we let $x = \frac{(t-n)\delta t}{a}$, then

$$\begin{aligned}\overline{\psi(x)} &= \pi^{-1/4} e^{-\frac{x^2}{2}} e^{-i\omega_0 x} \\ &= \pi^{-1/4} e^{-\frac{x^2}{2}} \left[\cos(\omega_0 x) - i \sin(\omega_0 x)\right].\end{aligned}$$

Therefore,

$$\Re[\overline{\psi(x)}] = \pi^{-1/4} e^{-\frac{x^2}{2}} \cos(\omega_0 x)$$

and

$$\Im[\overline{\psi(x)}] = -\pi^{-1/4} e^{-\frac{x^2}{2}} \sin(\omega_0 x)$$

Thus,

$$\sum_{x=\frac{-n\delta t}{a}}^{\frac{(N-1-n)\delta t}{a}} \Re\left[\overline{\psi(x)}\right]\Im\left[\overline{\psi(x)}\right] = -\pi^{-1/2}\sum_{x=\frac{-n\delta t}{a}}^{\frac{(N-1-n)\delta t}{a}} e^{-x^2}\cos(\omega_0 x)\sin(\omega_0 t)$$

$$= -\frac{\pi^{-1/2}}{2}\sum_{x=\frac{-n\delta t}{a}}^{\frac{(N-1-n)\delta t}{a}} e^{-x^2}\sin(2\omega_0 x)$$

It should be clear that, in general, this sum is not equal to zero. However, we note that

$$\lim_{\delta t\to 0}\frac{\delta t^2\sigma^2}{a}\sum_{t=0}^{N-1}\Re\left[\overline{\psi\left(\frac{(t-n)\delta t}{a}\right)}\right]\Im\left[\overline{\psi\left(\frac{(t-n)\delta t}{a}\right)}\right] = 0$$

□

Unfortunately, the next lemma also isn't true; like the previous lemma, however, it becomes approximately accurate when the sampling interval δt approaches zero.

LEMMA 4.4. Let $W(t) \in \ell^2(\mathbb{Z}_N)$ be a white noise signal with mean $\mu = 0$ and variance σ^2. Then

$$\mathbf{Var}[\Re[T_{n,a}[W]]] = \mathbf{Var}[\Im[T_{n,a}[W]]] = \frac{1}{2}\delta t\sigma^2.$$

PROOF. From Lemma 4.1, we know that $\mathbf{E}[\Re[T_{n,a}[W]]] = \mathbf{E}[\Im[T_{n,a}[W]]] = 0$. Therefore, we know that

$$\mathbf{Var}[\Re[T_{n,a}[W]]] = \mathbf{E}[(\Re[T_{n,a}[W]])^2]$$

and

$$\mathbf{Var}[\Im[T_{n,a}[W]]] = \mathbf{E}[(\Im[T_{n,a}[W]])^2].$$

We will only compute the real part; the proof of the imaginary part is very similar. First,

$$\begin{aligned}\mathbf{E}[(\Re[T_{n,a}[W]])^2] &= \mathbf{E}\left[\left(\sum_{t_1=0}^{N-1}\frac{\delta t}{\sqrt{a}}\Re\left[\overline{\psi\left(\frac{(t_1-n)\delta t}{a}\right)}\right]x(t_1)\right)\left(\sum_{t_2=0}^{N-1}\frac{\delta t}{\sqrt{a}}\Re\left[\overline{\psi\left(\frac{(t_2-n)\delta t}{a}\right)}\right]x(t_2)\right)\right]\\ &= \frac{\delta t^2}{a}\sum_{t=0}^{N-1}\mathbf{E}[W(t)^2]\Re^2\left[\overline{\psi\left(\frac{(t_2-n)\delta t}{a}\right)}\right].\end{aligned}$$

The last step follows from the linearity properties of the expected value operator and the fact that the $W(i)$'s are independent with mean $\mu = 0$. Therefore, for $i \neq j$, $\mathbf{E}[W(i)W(j)] = 0$. In addition, since $\mathbf{E}[W(i)] = 0$, $\mathbf{E}[W(i)^2] = \sigma^2$. We can rewrite the expression as

$$\mathbf{Var}[\Im[T_{n,a}[W]]] = \delta t\sigma^2\sum_{t=0}^{N-1}\frac{1}{a}\Re^2\left[\overline{\psi\left(\frac{(t-n)\delta t}{a}\right)}\right]\delta t.$$

In order for the lemma to be true, this expression will need to equal $\frac{1}{2}\delta t\sigma^2$. Unfortunately, it doesn't - it's only true when we let δt approach zero. The expression is actually a left Riemann sum. We approximate the sum with the corresponding integral; that is,

$$\begin{aligned}\lim_{\delta t\to 0}\sum_{t=0}^{N-1}\frac{1}{a}\Re\left[\overline{\psi\left(\frac{(t-n)\delta t}{a}\right)}\right]^2\delta t &= \int_{-\infty}^{\infty}\frac{1}{a}\Re\left[\overline{\psi\left(\frac{(t-n)}{a}\right)}\right]^2 dt\\ &= \int_{-\infty}^{\infty}\Re\left[\overline{\psi\left(u\right)}\right]^2 du\\ &= \int_{-\infty}^{\infty}\left[\pi^{-1/4}e^{-\frac{u^2}{2}}\cos(\omega_0 u)\right]^2 du.\end{aligned}$$

Integrating numerically, we find the integral equals 1/2. Of course, the result will not hold true for any fixed δt greater than zero. We have assumed that N is sufficiently large that we can approximate the summation with an integral with bounds $-\infty < t < \infty$[3]. Nevertheless, we grudgingly admit that

$$\mathbf{Var}[\Re[T_{n,a}[W]]] \approx \frac{1}{2}\delta t\sigma^2.$$

□

[3]Since wavelets are window functions, this is not actually a bad assumption. However, it will not hold true for the extreme edges of the wavelet transform.

We collect all of the previous lemmas to prove the following theorem (and later theorems). We assume that all of the lemmas are accurate. We will discuss potential problems with the assumptions we've made at the end of the chapter.

THEOREM 4.5. Let $W(t)$ be a white noise signal with mean $\mu = 0$ and variance σ^2. Then

$$\frac{|T_{n,a}[W]|^2}{\delta t \sigma^2/2} \sim \chi_2^2.$$

PROOF. This theorem is given in [**6**]. However, only an outline of the proof is given; we have filled in the gaps with the lemmas provided above.

(1) We want to show that both the real and imaginary parts of the wavelet transform of a white noise signal are normally distributed with mean $\mu = 0$; we proved this in Lemma 4.1.
(2) By dividing by their standard deviation - found in Lemma 4.4 - we can standardize both the real and the imaginary parts. That is,
$$\frac{\Re[T_{n,a}[W]]}{\sqrt{\delta t \sigma^2/2}} \sim N(0,1)$$
and
$$\frac{\Im[T_{n,a}[W]]}{\sqrt{\delta t \sigma^2/2}} \sim N(0,1).$$
(3) By Lemma 4.3, the real and imaginary parts of the Morlet wavelet transform of white noise are independent random variables.
(4) It is well known that the square of a random variable with a standard normal distribution has a chi-square distribution with one degree of freedom; furthermore, the sum of two *independent* random variables with chi-square distributions each with one degree of freedom has a chi-square distribution with two degrees of freedom. Thus,
$$\frac{|T_{n,a}[W]|^2}{\delta t \sigma^2/2} = \frac{\Re[T_{n,a}[W]]^2}{\delta t \sigma^2/2} + \frac{\Im[T_{n,a}[W]]^2}{\delta t \sigma^2/2} \sim \chi_2^2.$$

□

Note that our probability distribution does not depend on n or a — this is a remarkable fact, and it greatly reduces the work that we have to do.

We now introduce our significance test.

PROCEDURE 4. **One Signal Morlet Wavelet Significance Test**. Suppose $z \in \ell^2(\mathbb{Z}_N)$. We want to conduct a test to determine whether the Morlet wavelet transform coefficients of z are significantly different from the Morlet wavelet transform coefficients of a white noise signal $W \in \ell^2(\mathbb{Z}_N)$. We allow for some pre-specified α, the likelihood of a Type I error.

The null hypothesis of our test is that $\frac{|T_{n,a}[z']|^2}{\delta t/2} \sim \chi_2^2$, where z' is a normalized version of z that we will introduce momentarily; the alternative hypothesis is that $\frac{|T_{n,a}[z']|^2}{\delta t/2} \nsim \chi_1^2$. We reject or fail to reject the null hypothesis using the following steps.

(1) Normalize z using Procedure 1 from Chapter 1; we call this normalized signal z'.

(2) Compute the Morlet wavelet transform of z'.

(3) Compute the rejection region for the null hypothesis. That is, determine k such that $P\left(|T_{n,a}[z']|^2 \geqslant k\right) = \alpha$, or, equivalently, solve the following equation for k:

$$1 - \int_0^k \frac{1}{2\Gamma(1)} e^{-\frac{x}{2}} dx = \alpha.$$

(4) For all a and n, if $|T_{n,a}[z']|^2 \geqslant k$, reject the null hypothesis. Otherwise, we fail to reject the null hypothesis.

We provide examples of the Morlet wavelet significance test in Chapter 5.

1.2. Cross Spectrum Significance with the Morlet Wavelet. As in Chapter 3, we now introduce the Morlet wavelet cross spectrum significance test. We first introduce the relevant probability distribution; we will then construct the significance test. The following discussion is taken from [**7**].

DEFINITION 47. Let $x, y \in \ell^2(\mathbb{Z}_N)$. Then we define the **wavelet cross spectrum** of x and y by

$$C_{n,a}(x, y) = \overline{T_{n,a}[x]} T_{n,a}[y].$$

In general, we are more concerned with the square of the wavelet cross spectrum; that is,

$$|C_{n,a}(x,y)|^2 = |T_{n,a}[x]|^2 |T_{n,a}[y]|^2.$$

THEOREM 4.6. Let $W_1, W_2 \in \ell^2(\mathbb{Z}_N)$ be white noise signals with means $\mu_1 = \mu_2 = 0$ and variances $\sigma^2_{W_1}, \sigma^2_{W_2}$, respectively. Then

$$\frac{|C_{n,a}(W_1,W_2)|^2}{\sigma^2_{W_1}\sigma^2_{W_2}} \sim \frac{1}{4}\delta t^2 \kappa$$

where κ is a random variable with probability density function

$$f(x) = \frac{1}{2}K_0(x^{1/2})$$

where K_0 is the modified Bessel function of the second kind of order zero.

PROOF. Though the result seems intimidating, it is actually trivial to prove given past statistical results. Assuming the results of the last section hold true for small δt, we know

$$\frac{|T_{n,a}[W_1]|^2}{\delta t \sigma_x^2/2} \sim \chi_2^2$$

and

$$\frac{|T_{n,a}[W_2]|^2}{\delta t \sigma_y^2/2} \sim \chi_2^2.$$

From [**14**], we know that the product of two chi-square random variables each with two degrees of freedom is a random variable κ with probability density function

$$f(x) = \frac{1}{2}K_0(x^{1/2}).$$

Then

$$\frac{|C_{n,a}(W_1,W_2)|^2}{\sigma^2_{W_1}\sigma^2_{W_2}} = \frac{|T_{n,a}[W_1]|^2}{\sigma^2_{W_1}} + \frac{|T_{n,a}[W_2]|^2}{\sigma^2_{W_2}} \sim \frac{1}{4}\delta t^2 \kappa$$

□

We now introduce our last significance test.

Procedure 5. **Cross Spectrum Morlet Wavelet Significance Test**. Suppose $z_1, z_2 \in \ell^2(\mathbb{Z}_N)$. We want to conduct a test to determine whether the cross spectrum coefficients of z_1 and z_2 are significantly different from the cross spectrum Morlet wavelet transform coefficients of two white noise signals $W_1, W_2 \in \ell^2(\mathbb{Z}_N)$. We allow for some pre-specified α, the likelihood of a Type I error.

The null hypothesis of our test is that $\frac{|C_{n,a}(z_1', z_2')|^2}{\delta t^2/4} \sim \kappa$, where z' is a normalized version of z that we will introduce momentarily; the alternative hypothesis is that $\frac{|C_{n,a}(z_1', z_2')|^2}{\delta t^2/4} \nsim \chi_1^2$. We reject or fail to reject the null hypothesis using the following steps.

(1) Normalize z_1 and z_2 using Procedure 1 from Chapter 1; we call these normalized signals z_1' and z_2'.
(2) Compute the Morlet wavelet transforms of z_1' and z_2'.
(3) Compute the rejection region for the null hypothesis. That is, determine k such that $P\left(\frac{|C_{n,a}(z_1', z_2')|^2}{\delta t^2/4} \geqslant k\right) = \alpha$, or, equivalently, solve the following equation for k:

$$1 - \int_0^k \frac{1}{2} K_0(x^{1/2}) dx = \alpha.$$

(4) For all n and a, if $\frac{|C_{n,a}(z_1', z_2')|^2}{\delta t^2/4} \geqslant k$, reject the null hypothesis. Otherwise, we fail to reject the null hypothesis.

We provide examples of the Morlet wavelet cross spectrum significance test in Chapter 5.

CHAPTER 5

Applications

1. Sample Significance Tests

We provide a few sample significance tests. The data sets are provided by the *Gloeo* group at Dartmouth. Unfortunately, no pair of these sample data sets was suitable for the Morlet wavelet cross spectrum test; instead, we used the famous Canadian hare and lynx population data set for our test.

1.1. Sample Haar Wavelet Significance Tests. Many of the sample data sets provided by the *Gloeo* group contained fewer than 20 data points; I felt that an analysis of such a short data set using the Morlet wavelet would be corrupted by edge effects. Though I did not discuss them in technical detail, the reader should know that edge effects are the result of a wavelet exceeding the boundaries of the signal of interest; if it gives the reader any intuition about how these edge effects might occur, consider that highly dilated wavelets (those with scale term a close to zero) produce the worst edge effects.

I designed both of the Haar wavelet significance tests with the hope that I could more adequately analyze these short signals—the Haar wavelet is the shortest of the discrete wavelets, and does not provoke as harsh edge effects as other wavelets. Unfortunately, the Haar wavelet does not have many other desirable properties.

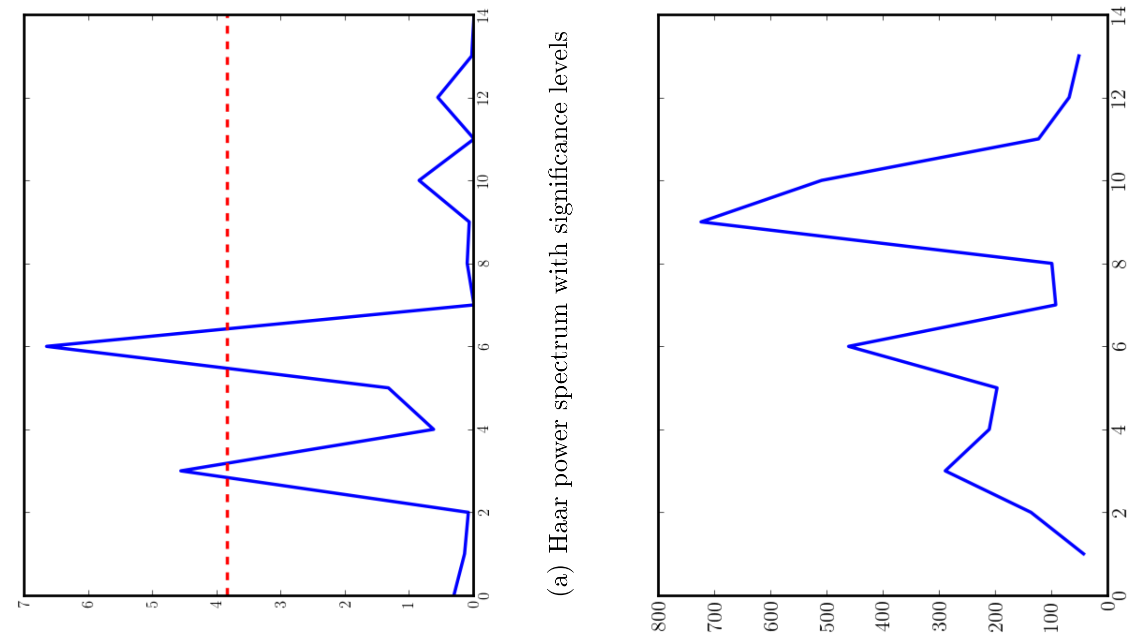
As an orthonormal decomposition, the discrete Haar wavelet transform suffers from the same kinds of leakage problems as the discrete Fourier transform. In addition, it can be very hard to interpret Haar transform coefficients—they are not easily translated into units of time or frequency. Nevertheless, I believe the Haar significance tests offer a viable alternative to the Morlet wavelet significance tests.

In the future, I hope to eschew interpreting the Haar wavelet coefficients directly; instead, I hope to use the statistically significant coefficients to build models of ecological situations.

In theory, statistically significant coefficients reflect the underlying system that produced the signal—a model built using these coefficients should have some predictive validity. In addition, I would like to consider how significant cross spectrum coefficients could validate an ecological model.

Figures 5.1 and 5.2 plot *Gloeo* recruitment data per m^3 for 2008 and 2009 at Lake Sunapee in New Hampshire; Figure 5.3 plots the cross spectrum of the 2008 and 2009 data sets. Remember that the Haar wavelet transform coefficients have been concatenated into a single vector.

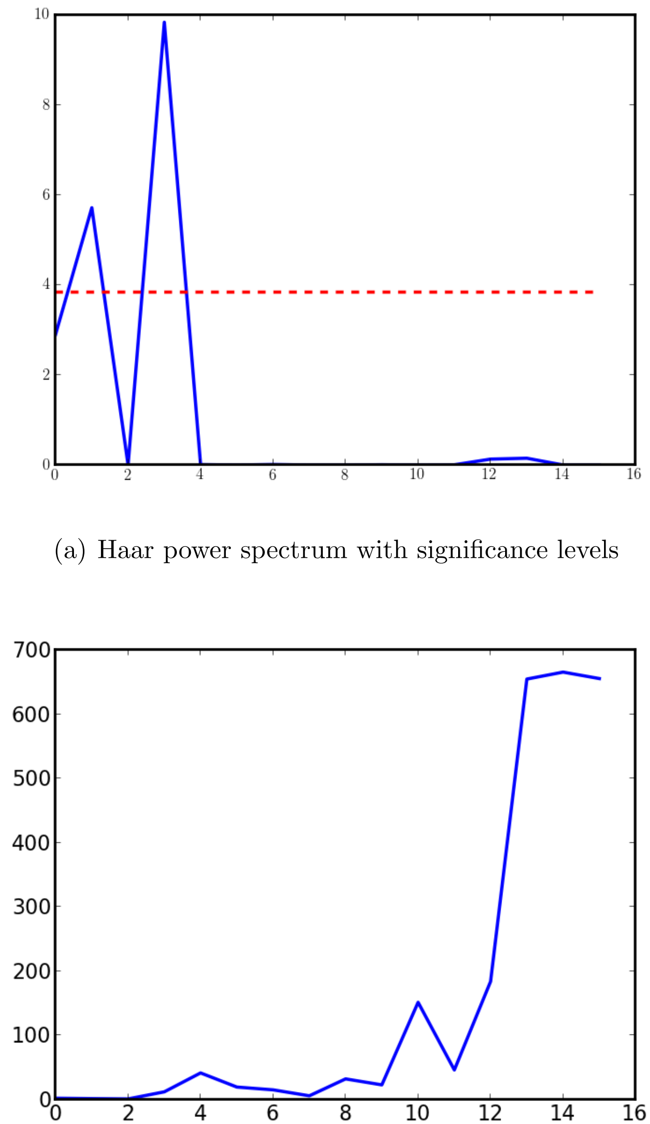
Importantly, these plots indicate statistically significant cross spectrum coefficients; these coefficients indicate some periodic behavior for both summers on the two to three week scale.



(a) Haar power spectrum with significance levels

(b) Time-amplitude plot

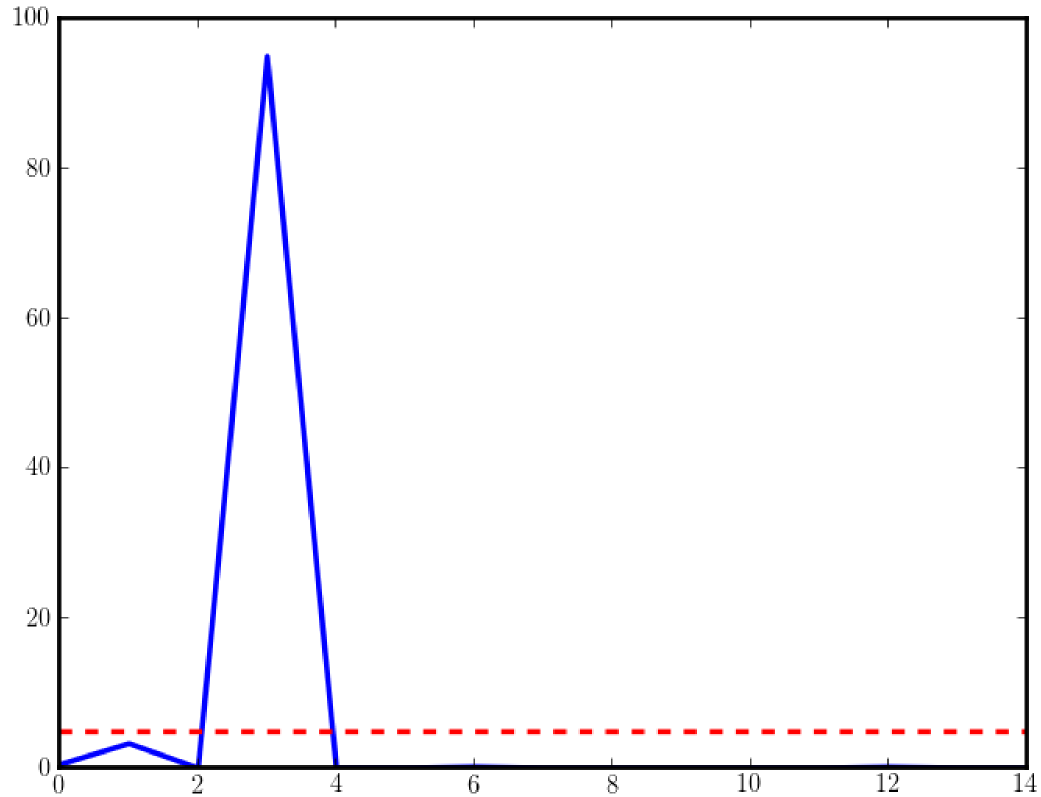
FIGURE 5.1. Plots of weekly *Gloeo* recruitment per m^3 for 6/08 - 9/08. Red line indicates 95% confidence level.



(a) Haar power spectrum with significance levels

(b) Time-amplitude plot

FIGURE 5.2. Plots of weekly *Gloeo* recruitment per m^3 for 6/09 - 9/09. Red line indicates 95% confidence level.



(a) Haar cross spectrum with significance levels

FIGURE 5.3. Cross spectrum for weekly *Gloeo* recruitment per m^3 for 2008 and 2009. Red line indicates 95% confidence level.

1.2. Morlet Wavelet Significance Tests. Many sources, including [**13**], guard against an apparently common "style over substance" charge brought against the wavelet transform: that the continuous wavelet transform produces pretty pictures, but cannot generate useful quantitative results. These critiques inspired the introduction of the first Monte Carlo simulations used to produce statistical significance tests; in turn, the lack of statistical rigor in these tests inspired the probabilistic derivations found in [**6**], [**7**], and this thesis.

Despite this apparent introduction of rigor, the Morlet wavelet transform still has many flaws. For example, consider Figure 5.4: despite testing at the 95% confidence level, the region of significance remains a diffuse and hard-to-interpret band within the scaleogram. It's hard to know how a researcher could use this kind of data to prove a result or to promote future research—the transform just isn't accurate enough. For example, if we wanted to build a model using the significance test found in Figure 5.4, which transform coefficients would we base our model upon?

We would construct a model using the most statistically significant coefficients; unfortunately, since different components of a signal can have different relative strengths, this could exclude possibly vital coefficients from the model. For example, an attempt to construct a model of the signal

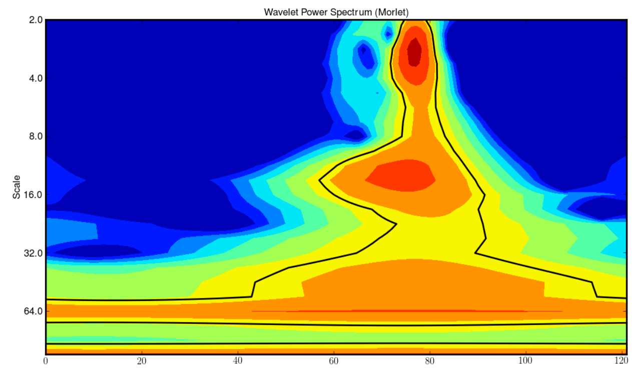
$$f(x) = \frac{\sin(2\pi x)}{4} + \sin(4\pi x)$$

using this method would inevitably exclude the $\frac{\sin(2\pi x)}{4}$ term. Without sophisticated *a priori* knowledge of the system being examined by the transform, it's hard to know how to use the Morlet wavelet transform effectively.

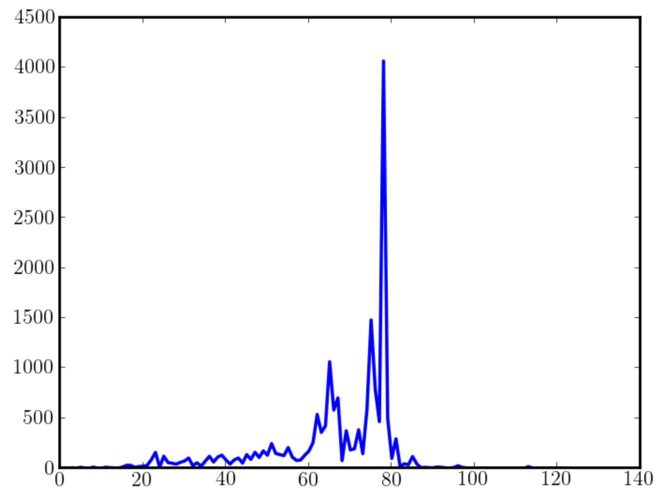
Figure 5.5 shows plots of the light intensity at Lake Sunapee recorded every ten minutes for 2008. When converted from ten minute intervals to hours, the two narrow bands of significance running across the center of the scaleogram correspond to periodic behavior on 12 and 24 hour intervals—this is what we should expect. Most interestingly, the transform contains patterns of statistically significant coefficients that indicate daily or weekly periodic behavior during parts of spring, summer, and fall. Again, without a well-reasoned procedure for handling the data, it's unclear what a researcher could do with this information.

Figure 5.6 shows the cross spectrum test of the classic Canadian hare-lynx population data. The cross spectrum indicates significant correlation in periodicity on the four to seven year scale between the hare and lynx population levels for about thirty years.

The hare-lynx data set was taken from `http://dave-reed.com/csc121.F10/Labs/Lab1/`.

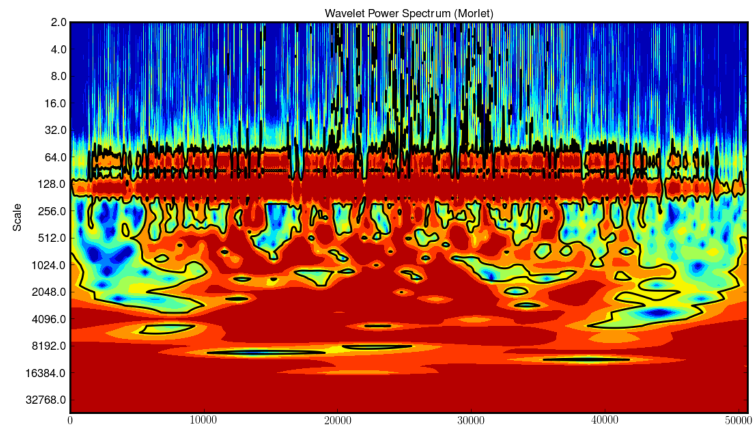


(a) Morlet power spectrum with significance levels

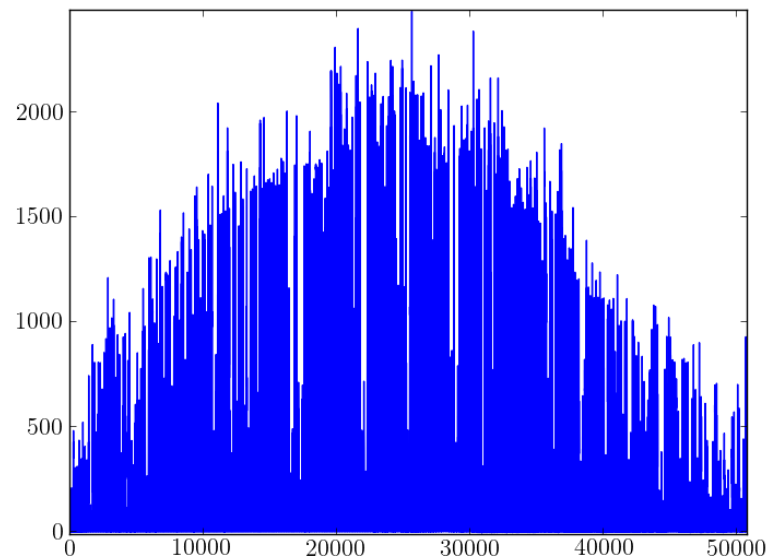


(b) Time-amplitude plot

FIGURE 5.4. Number of *Gloeo* colonies for 5/5/10 - 10/6/10; recorded daily. Black line indicates 95% confidence level.

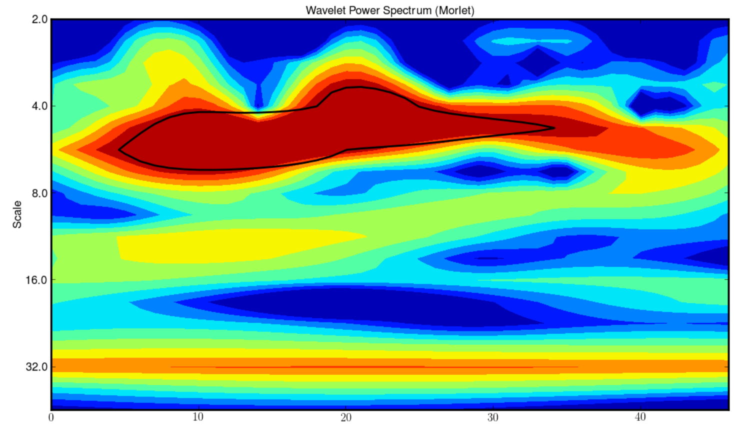


(a) Morlet power spectrum with significance levels

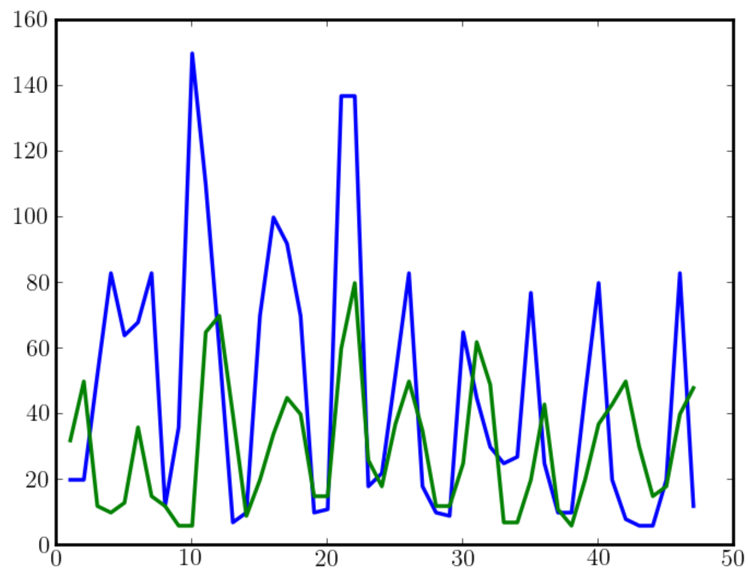


(b) Time-amplitude plot

FIGURE 5.5. Light intensity in ten minute intervals for 2008. Black line indicates 95% confidence level.



(a) Morlet cross spectrum with significance levels



(b) Blue line indicates hare data; green line indicates lynx.

FIGURE 5.6. Hare - lynx population data in thousands for 1845-1937. Black line indicates 95% confidence level.

2. Conclusion

2.1. Shortcomings of Significance Tests. In the past several sections, I indicated problems with specific significance tests. However, I would like to note a serious problem with all of the significance tests in this thesis.

In Chapter 1, I defined a Type II error—the error associated with failing to reject the null hypothesis even though it is false. All of the significance tests defined in this thesis avoid Type I errors by constructing appropriate rejection regions based upon the null hypothesis in question. However, none of the tests account for Type II errors; this is because it is much harder to compute the probability of Type II errors. For example, for the one signal Haar significance test, the probability of committing a Type II error is the probability that a random variable that is *not* chi-square distributed with one degree of freedom registers values that a chi-square random variable with one degree of freedom registers 95% of the time.

When we don't *a priori* know anything about a signal, we don't have any means of computing the probability of committing a Type II error—at least not that I can see. In fact, to me, this seems like a baffling and generally unsolvable problem. This is a shame, because without knowing the probability of a Type II error, we don't know how much useful information we're throwing out when we conduct a significance test.

2.2. Future Research. I mentioned in the introduction that continuous wavelet transforms are best suited for ecological research—I said this because the Morlet wavelet transform coefficients are easily interpreted in terms of frequency and position. However, I'm convinced that there is not very much value in examining a wavelet transform scaleogram by itself—I outlined my critiques of the Morlet wavelet transform in the sections above.

Instead, I believe the future of wavelets in ecological research will be the use of significance tests on discrete wavelet transforms to construct models. To advance these aims, I would like to extend the results I found for the Haar wavelet transform of white noise to more complex discrete wavelet transforms—the Shannon wavelet, for example, or the Daubechies wavelet, which we did not discuss. Deriving probability distributions of wavelet transforms of white noise for these wavelets should be relatively straightforward, and these wavelets have many desirable properties relative to the Haar wavelet, including improved frequency localization.

Ecology can benefit from wavelets—ecological data sets, like the data sets examined above, often exhibit the kind of short-term periodic behavior best analyzed by wavelets. Nevertheless, the methods need to become more sophisticated in order for researchers to make valuable and well-justified claims about data sets.

Appendix. Python Implementation

2.3. Python Code for Haar Wavelet Transform. The following code computes the complete Haar wavelet transform; in addition to Python 2, the code requires the Python package `numpy`. The code is adapted from pseudocode found in [**11**].

```
import numpy as np

def haar_decomp(my_signal):
    my_signal = my_signal/np.sqrt(len(my_signal))
    my_h = len(my_signal)
    def decomp_step(signal,h):
        haar_coef = np.zeros(h)
        for i in range(0,h/2):
            haar_coef[i] = (signal[2*i] + signal[2*i+1])/np.sqrt(2)
            haar_coef[h/2 + i] = (signal[2*i] -signal[2*i+1])/np.sqrt(2)
        signal[0:h] = haar_coef
        return signal

    while my_h>1:
        my_signal = decomp_step(my_signal,my_h)
        my_h = my_h/2

    return my_signal
```

The following code reconstructs a signal from Haar coefficients; that is, it is the inverse of the code provided above. The code is my own.

```
def haar_recon(my_haar):
    my_h = 1
    len_haar = len(my_haar)
    count = 0
    def recon_step(haar_coef,h):
        recon = np.zeros(2*h)
        for i in range(0,h):
            recon[2*i] = (haar_coef[i]+haar_coef[i+h]*np.power(2,count/2.0))
            recon[2*i+1] = (haar_coef[i]-haar_coef[i+h]*np.power(2,count/2.0))

        haar_coef[0:2*h] = recon
        return haar_coef

    while my_h < len(my_haar):
        my_haar = recon_step(my_haar,my_h)
        my_h = my_h*2
        count +=1.0
    return my_haar
```

2.4. Python Code for Hanning Filter.

```
from numpy import zeros, loadtxt
from numpy.fft import fft,ifft
from pylab import *
```

```
fname = 'my_filename.dat'
my_z = loadtxt(fname)



def hanning_filter(z):
    length_z = len(z)
    H_f = zeros(length_z)
    H_f[0] = .5
    H_f[1] = -.25
    H_f[-1] = -.25
    fft_z = fft(z)
    H_f_z = ifft(fft(H_f)*fft(fft(z)))
    fft_ab = fft(z)
    return H_f_z, fft_ab, fft(H_f)

h,new_z,h_f = hanning_filter(my_z)

subplot(2,2,1)
plot(x,my_z, linewidth=2.5)

subplot(2,2,2)
plot(abs(h_f),linewidth=2.5 )

subplot(2,2,3)
plot(abs(h)**2,linewidth=2.5)
```

```
subplot(2,2,4)
plot(abs(new_z) ** 2,linewidth=2.5)

show()
```

2.5. Python Code Demonstrating Morlet Wavelet Transform Real and Imaginary Part Non-Independence.

```
    from numpy import *

my_x = linspace(0,10,1000)
my_dx = .01
my_n = 5
my_a = 1

def morlet_test(x,dx,n,a):
    i = 1j
    t = ((x - 2*ones(len(x)))*dx)/a
    morlet = (pi ** (-.25)) * exp(-(t ** 2)/2.0) * exp(-i*6.0*t)
    my_test = ((dx ** 2)/a)*sum(real(morlet)*imag(morlet))
    print my_test


print morlet_test(my_x,my_dx,my_n, my_a)
```

2.6. Python Code for One Signal Haar Significance.

```
import numpy as np
import pylab
from scipy.optimize import newton
from scipy.stats import chi2
import pywt

fname = 'gloeo2009.dat'
signal = np.loadtxt(fname)

signal = (signal - np.mean(signal))/np.std(signal)
sig_level = .95

db1 = pywt.Wavelet('db1')



coeffs = pywt.wavedec(signal, db1, mode = 'per')
my_coeffs = np.concatenate(coeffs)



def find_k(significance_level):
    def chi_cdf(my_k):
        return chi2.cdf(my_k,1) - significance_level
    good_k = newton(chi_cdf,significance_level)
    return good_k
```

```
k = find_k(sig_level)

x = range(len(my_coeffs))

pylab.plot(x,np.power(my_coeffs,2),linewidth = 2.5)
pylab.plot(x, k*np.ones(len(x)), color="red",  linewidth=2.5, linestyle="--")
pylab.show()
```

2.7. Python Code for Cross Spectrum Haar Significance.

```
import numpy as np
import pylab
from scipy.integrate import quad
from scipy.optimize import newton
from scipy.special import k0
import pywt

fname_1 = 'gloeo2008.dat'
signal_1 = np.loadtxt(fname_1)
signal_1 = (signal_1 - np.mean(signal_1))/np.std(signal_1)

fname_2 = 'gloeo2009.dat'
signal_2 = np.loadtxt(fname_2)
signal_2 = (signal_2 - np.mean(signal_2))/np.std(signal_2)

sig_level = .95

db1 = pywt.Wavelet('db1')
```

```
coeffs_1 = pywt.wavedec(signal_1, db1, mode = 'per')
my_coeffs_1 = np.concatenate(coeffs_1)

coeffs_2 = pywt.wavedec(signal_2, db1, mode = 'per')
my_coeffs_2 = np.concatenate(coeffs_2)



def find_k(significance_level):
    def bessel_cdf(k):
        result,err = quad(lambda x: (1/(np.pi*np.sqrt(x)))*k0(np.sqrt(x)), 0, k)
        return result - significance_level

    good_k = newton(bessel_cdf,5*significance_level)
    return good_k



k = find_k(sig_level)

x = range(len(my_coeffs_1))

pylab.plot(x,np.power(my_coeffs_1,2)*np.power(my_coeffs_2,2),linewidth = 2.5)
pylab.plot(x, k*np.ones(len(x)), color="red",  linewidth=2.5, linestyle="--")
pylab.show()
```

All plots in this thesis were generated using the Python package `Matplotlib`. The continuous Morlet wavelet transforms and one-signal significance tests were computed using the Python package `KPyWavelet`. The cross spectra plots and significance tests were computed by adapting the `KPyWavelet` source code—this adapted code was not provided but is available upon request.

Bibliography

[1] C. Bingham, M.D. Godfrey, and J.W. Tukey, Modern Techniques of Power Spectrum Estimation. *IEEE Transactions On Audio and Electroacoustics* **15(2)** (1967) 56-66.

[2] B. Cazelles et al., Wavelet Analysis of Ecological Time Series. *Oecologia* **156(2)** (2008), 287-304.

[3] C. Chui, *An Introduction to Wavelets*, Academic Press, Boston,1992.

[4] M. Frazier, *An Introduction to Wavelets Through Linear Algebra*, Springer, New York, 1999.

[5] J. Freund, *Mathematical Statistics with Applications*, Pearson, New Jersey, 2004.

[6] Z. Ge, Significance Tests for the Wavelet Power and the Wavelet Power Spectrum. *Ann. Geophys.* **25** (2007) 2259-2269.

[7] Z. Ge, Significance Tests for the Wavelet Cross Spectrum and Wavelet Linear Coherence. *Ann. Geophys.* **26** (2008) 3819-3829.

[8] D. Maraun and J. Kurths, Cross Wavelet Analysis: Significance Testing and Pitfalls. *Nonlinear Processes in Geophysics* **11** (2004) 505-514.

[9] E. Lukacs and E. King, A Property of the Normal Distribution. *Ann. Math Statist.* **25 (2)** (1954), 389-394.

[10] Redding, Nicholas J. Estimating the parameters of the K distribution in the intensity domain. No. DSTO-TR-0839. Electronics Research Lab. Salisbury, Australia, 1999.

[11] E. Stollnitz, T. DeRose, and D. Salesin, Wavelets for Computer Graphics: A Primer, Part 1. *IEEE Computer Graphics and Applications* **15(3)** (1995), 76-84.

[12] E. Stollnitz, T. DeRose, and D. Salesin, Wavelets for Computer Graphics: A Primer, Part 2. *IEEE Computer Graphics and Applications* **15(4)** (1995), 75-85.

[13] C. Torrence and G. Compo, A Practical Guide to Wavelet Analysis. *Bulletin of the American Meterological Society* **79(1)** (1998), 61-78.

[14] W.T. Wells, R.L. Anderson, and J.W. Cell, The distribution of the product of two central or non-central chi-square random variates. *Ann. Math. Stat.* **33** (1962) 1016-1020.