

# Wavelets

Shane Scott

Kansas State University Honors Program

2012

### **Abstract**

Fourier spectral analysis provides methods of signal compression and approximation, but presents some limitations to practical implementations. The Fourier transform of a signal may be thought of as its representation over a (pseudo) basis of complex exponentials, which while localized in frequency space, are delocalized in time. In contrast with the Fourier basis, a wavelet basis can be both frequency and time localized. Such a basis allows transformations between representations to be calculated locally, and partial representations can often be calculated even when only a portion of the signal is known. We demonstrate and compare low-loss signal compression using both methods. Through a discussion of the advantages of wavelet representations we will motivate a theoretical discussion of alternative bases and give proofs of their construction and in the space of square summable complex sequences and the space of square Lebesgue integrable functions.

# Contents

<b>1</b>	<b>Introduction: Wavelet Compression of a Digital Signal</b>	<b>2</b>
<b>2</b>	<b>Discrete Wavelets</b>	<b>16</b>
<b>3</b>	<b>Continuous Wavelets</b>	<b>23</b>
<b>4</b>	<b>Appendix</b>	<b>29</b>
4.1	Lebesgue Measure and Integration . . . . .	29
4.2	Hilbert Spaces . . . . .	32
4.3	$L^2(\mathbb{R})$ as a Hilbert Space . . . . .	33

# List of Figures

1.1	A Walken Signal $z$ . . . . .	3
1.2	The Walken Spectrum $ \hat{z} $ . . . . .	4
1.3	The 90% Compressed Walken Spectrum $ \hat{z} $ . . . . .	5
1.4	The 90% Compressed Walken Signal $z$ . . . . .	6
1.5	The 95% Compressed Walken Spectrum $ \hat{z} $ . . . . .	6
1.6	The 95% Compressed Walken Signal $z$ . . . . .	7
1.7	Christopher Walken on SNL . . . . .	7
1.8	The Mexican Hat wavelet in standard space . . . . .	8
1.9	The Mexican Hat wavelet in frequency space . . . . .	8
1.10	The D4 wavelet in space . . . . .	9
1.11	The D4 wavelet in frequency space . . . . .	9
1.12	1st Stage Wavelet Representation $z_{(1)}$ . . . . .	12
1.13	2st Stage Wavelet Representation $z_{(2)}$ . . . . .	12
1.14	3rd Stage Wavelet Representation $z_{(3)}$ . . . . .	13
1.15	4th Stage Wavelet Representation $z_{(4)}$ . . . . .	13
1.16	90% Compressed 4th Stage Wavelet Representation $z_{(4)}$ . . . . .	13
1.17	90% Wavelet Compressed $z$ . . . . .	14
1.18	95% Compressed 5th Stage Wavelet Representation $z_{(5)}$ . . . . .	14
1.19	95% Compressed $z_{(5)}$ . . . . .	14
1.20	Comparison % Error by Compression Ratio . . . . .	15

## Acknowledgement

This work was supported by a scholarship from the Center for the Integration of Undergraduate, Graduate, and Postdoctoral Research of the Mathematics Department at Kansas State University under the direction of Dr. Virginia Naibo and in collaboration with Brian Moore and Vincent Pigno.



## Notation

$\mathbb{Z}$	the set of integers
$\mathbb{Z}^+$	the set of positive integers
$\mathbb{N}$	the set of non-negative integers
$\mathbb{R}$	the set of real numbers
$\mathbb{R}^+$	the set of positive real numbers
$\mathbb{R}_0^+$	the set of non-negative real numbers
$\mathbb{C}$	the set of complex numbers
$\mathcal{P}(A)$	set of all subsets of $A$
$A^c$	set compliment of $A$
$\ell^1(\mathbb{Z})$	the absolutely of summable sequences
$\ell^2(\mathbb{Z})$	the set of square summable sequences
$\ell^2(\mathbb{Z}_N)$	$N$ length complex vectors
$L^2(A)$	the set of square integrable complex valued functions on $A$
$\text{clos}_d A$	$d$ induced metric topology closure of $A$
$\sum_{a \in A} z(a)$	the summation of $z$ over set $A$
$\prod_{a \in A} z(a)$	the product of $z$ over set $A$
$\int_A f(x) dx$	the Lebesgue integral of $f$ over $A$
$\langle v w \rangle$	inner product of $v$ and $w$
$v * w$	convolution of $v$ and $w$
$v \perp w$	$v$ orthogonal to $w$
$V \oplus W$	orthogonal direct sum of $V$ and $W$
$ z $	modulus of $z$
$\ z\ $	norm of $z$
$\bar{z}$	complex conjugate of $z$
$\tilde{z}$	conjugate reflection of $z$
$R_n z$	$n^{th}$ translation of $z$
$\hat{z}$	Fourier or Discrete Fourier transform of $z$
$\check{f}$	inverse Fourier transform of $f$
$f : A \rightarrow B$	$f$ is a set function from $A$ to $B$
$f : a \mapsto b$	$f$ maps $a$ to $b$
$\delta$	Dirac delta sequence
$I_n$	$n \times n$ identity matrix

# Chapter 1

## Introduction: Wavelet Compression of a Digital Signal

Informally wavelets are vectors in a space whose translations and dilations span the space. Often wavelets may be chosen which are localized in both their temporal and frequency representations. Wavelets are of interest both to theory and application, so it may be best to demonstrate by an example. The following informal presentation of wavelet applications to data compression techniques may help to motivate the theoretical discussion of their existence and construction that follows.

In an increasingly data-filled age where few are unfamiliar with concepts of streaming or data transfer, more and more are routinely initiated merely from internet exposure, even while the underlying mathematics goes unrecognized. Efficiency of such data transmission is of great concern with the current proliferation of sites streaming music, video, photos. Signals (such as electrical and optical signals speeding through cables and antennas across the country) are best modeled as vectors in the space:

$$L^2(\mathbb{R}) = \left\{ v : \mathbb{R} \rightarrow \mathbb{C} \mid \int_{\mathbb{R}} |v(t)|^2 dt < \infty \right\}.$$

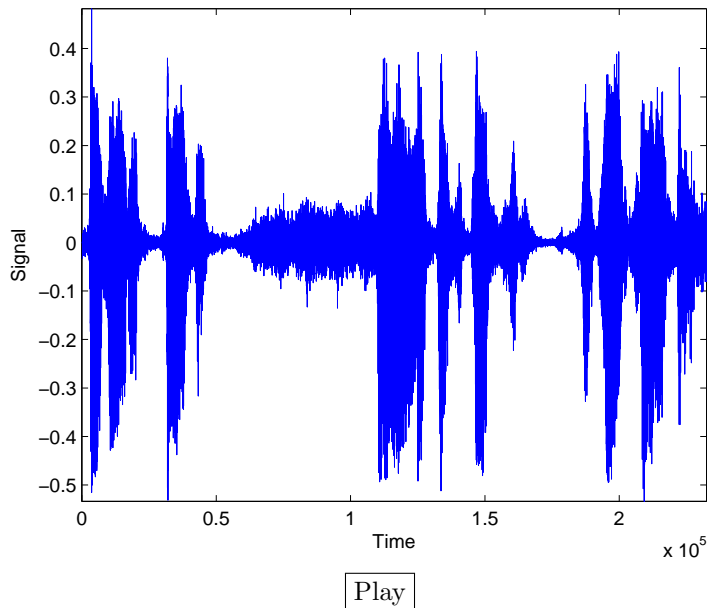
In the digital landscape, the discrete versions of signals instead lie in

$$\ell^2(\mathbb{Z}) = \left\{ v : \mathbb{Z} \rightarrow \mathbb{C} \mid \sum_{n \in \mathbb{Z}} |v(n)|^2 < \infty \right\},$$

or in the practice of computation, we may have discrete, finite length vectors

$$\ell^2(\mathbb{Z}_N) = \{v : \mathbb{Z}_N \rightarrow \mathbb{C}\}.$$

Figure 1.1: A Walken Signal  $z$



These are simply  $N$  length vectors of complex values. We will develop some of the theory of wavelets in these spaces. But for now, let's consider the signal  $z \in \ell^2(\mathbb{Z}_N)$  shown in Figure 1.1.

The signal  $z$  represents a digital sound signal of Christopher Walken performing a line from an April 8, 2000, *Saturday Night Live* sketch. (Press the play button in Figure 1.1 to listen to the clip.) Suppose we are faced with the problem of storing a great many such signals in a database, or transmitting the signal as fast as possible. Lossy compression techniques to such problems look to find ways of representing signals (or at least an approximation) with as little information as possible. A fairly simple method of digital compression is Fourier compression.

The discrete fourier transform of  $z$  is  $\hat{z}$ , also an  $N$  length vector, given by the formula

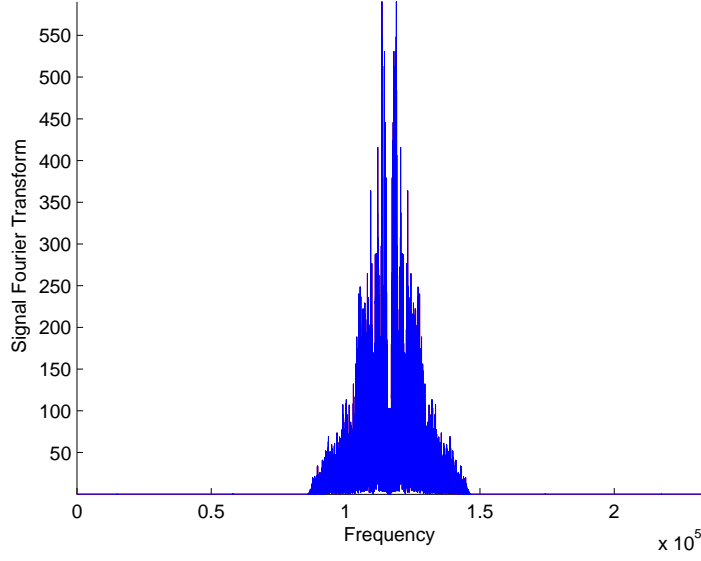
$$\hat{z}(m) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} z(n) e^{-2\pi i m n / N}. \quad (1.1)$$

This transformation is reversed by the inverse transformation

$$z(n) = \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} \hat{z}(m) e^{2\pi i m n / N} = \left\langle \frac{e^{-i2\pi m n / N}}{\sqrt{N}} \middle| \hat{z} \right\rangle, \quad (1.2)$$

and we can then think of  $\hat{z}$  as the representation of  $z$  over a basis of complex

Figure 1.2: The Walken Spectrum  $|\hat{z}|$



exponentials (essentially sine waves):

$$z(n) = \sum_{m=0}^{N-1} \left\langle \frac{e^{i2\pi mn/N}}{\sqrt{N}} \middle| z \right\rangle \frac{e^{i2\pi mn/N}}{\sqrt{N}},$$

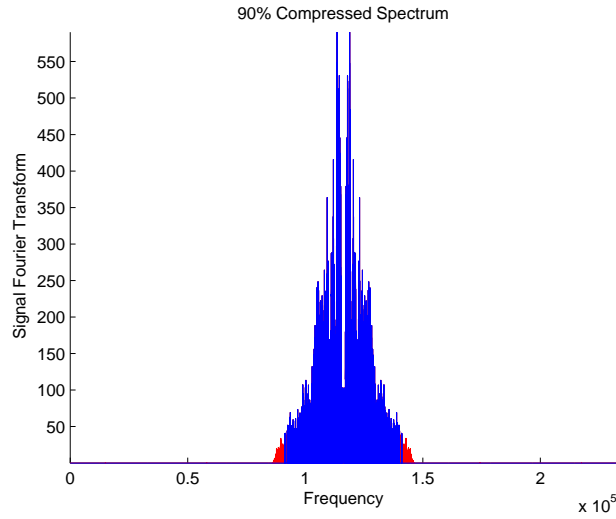
while  $z$  is its own representation over the standard basis. In this sense  $z$  is the *temporal* or *spatial* representation while  $\hat{z}$  is the *frequency* space representation of  $z$ . The component-wise absolute value of  $\hat{z}$  is called the *spectrum* of  $z$ . The spectrum of Mr. Walken's line is shown in Figure 1.2.

Notice that a small number of positions of  $\hat{z}$  have a large value, while many have values close to zero. Thus we can discard many of these position and replace them only with zero, we can perform the inverse Fourier transform to obtain a very close approximation of  $z$ . By transmitting only the positions of  $\hat{z}$  above some threshold value, we have much less data to transmit, but can still reconstruct a very close approximation to  $z$ . The Fourier compression algorithm at this point goes as follows. To compress and send a  $p$  percent sized representation of  $z$ :

1. Calculate  $\hat{z}$ .
2. Order the values of  $\hat{z}$  by their values in modulus and calculate  $a$ , their  $p^{th}$  percentile.
3. If a component of  $\hat{z}$  is less than  $a$  in absolute value, discard it and replace it with 0.



Figure 1.3: The 90% Compressed Walken Spectrum  $|\hat{z}|$



4. Transmit the positions and values of the non-zero positions of  $\hat{z}$ .
5. Reconstruct the approximation by taking the inverse Fourier transform.

Figures 1.3 and 1.5 show the 90% and 95% compressed spectrum of  $z$  in blue while the original is shown for comparison in red. (The thickness of the graph lines in the figures obscures the sparsity of the non-zero coefficients post-compression, but only 10%/5% remain.) Figures 1.4 and 1.6 show the reconstructed approximation to  $z$  in blue with the original  $z$  in red. Press the "Play" button by their graph to listen. Thus we can reduce by 95% or more the amount of data describing  $z$  and still have a very close approximation to Mr. Walken's line.

This technique is good, but still has a number of drawbacks. Note first that in the compressed versions of  $z$  as we increase the compression ratio the signal to background ratio is getting much worse, that is most of the information loss occurs on the most important parts of the signal. But for modern applications like web streaming, this method has an even bigger problem. Note that the inversion formula for  $z$  from  $\hat{z}$  (Eq. 1.2) each component of  $z$  must be calculated using every component of  $\hat{z}$ , so the inversion calculation cannot begin until all of  $\hat{z}$  is transmitted, which is decidedly not how web streaming operates. We call this problem the time *delocalization* of the complex exponential basis. Without going into detail, we can say compression of video signals works much the same way. Take a peak at this video clip of Mr. Walken delivering his line (Figure 1.7).

Although this example of Mr. Walken is a two dimensional example, the two dimensional case is very similar to that of the one dimensional; a Fourier transform exists and the coefficients depend on every spatial pixel, that is, they

Figure 1.4: The 90% Compressed Walken Signal  $z$

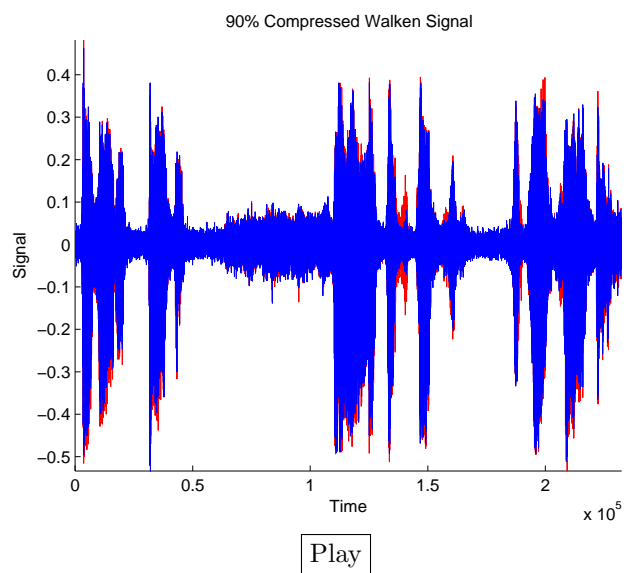


Figure 1.5: The 95% Compressed Walken Spectrum  $|\hat{z}|$

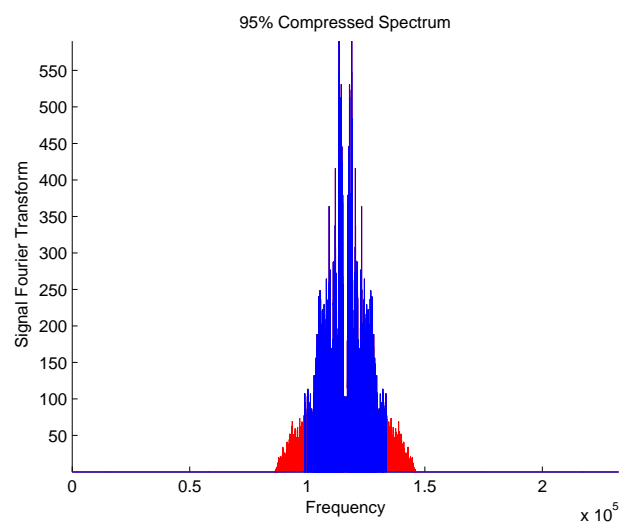
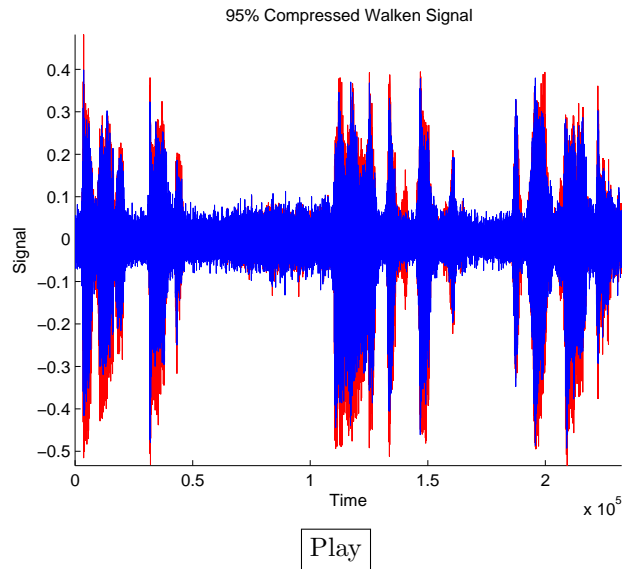


Figure 1.6: The 95% Compressed Walken Signal  $z$



are not spatially localized. Observe that the background behind Mr. Walken changes very little throughout the clip. A better algorithm might avoid sending redundant information to update the background, but because the values of  $\hat{z}$  depend on all components of  $z$ , (see Eq. 1.1), we cannot avoid sending some of this information. That is, the standard basis which expresses the time representation  $z$  is delocalized in the frequency space. But a modification of the compression algorithm can overcome these drawbacks.

The compression algorithm we have expressed depends only on having a basis over which the representation of  $z$  has as many zeros or small values, so

Figure 1.7: Christopher Walken on SNL



Figure 1.8: The Mexican Hat wavelet in standard space

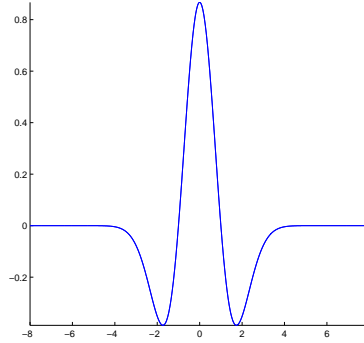
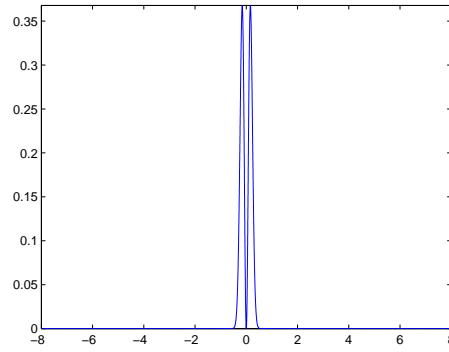


Figure 1.9: The Mexican Hat wavelet in frequency space

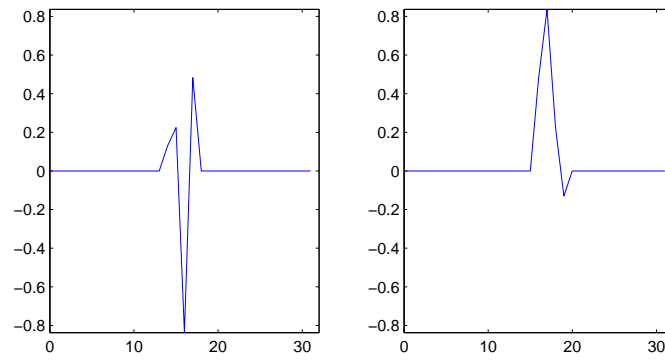


instead we can hope to find a basis which is localized both in the standard space and in the frequency space. Qualitatively this means that we will find a basis of vectors which have only a small section of non-zero components and whose Fourier transforms have only a small section of non-zero coefficients. One example is the Mexican Hat wavelet, whose time and frequency representations are shown in Figures 1.8 and 1.9. A basis may be formed via their spatial translations, which will avoid the localization objection raised to the complex exponential basis. A detailed technical discussion of this may be found in the succeeding chapters.

For discrete finite length vectors we form a first stage wavelet basis by taking the even spatial translations of two vectors  $u$  and  $v$ . We will choose  $u$  such that only low frequency components are non-zero, while  $v$  will have high frequency components non-zero. Figure 1.10 shows one such example, Daubechies's D4 wavelet filter pair.

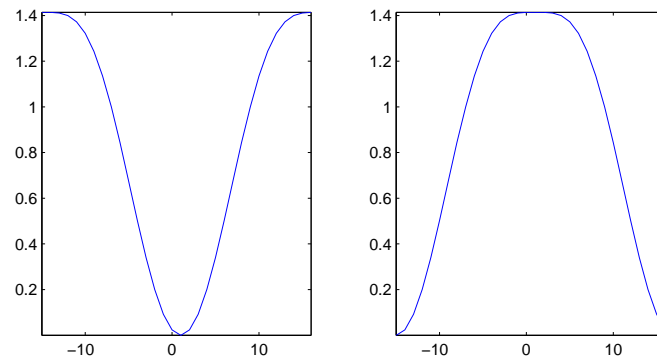
Then the components of  $z$  represented over the translations of  $u$  will con-

Figure 1.10: The D4 wavelet in space



(Left) High frequency filter. (Right) Low frequency filter.

Figure 1.11: The D4 wavelet in frequency space



(Left) High frequency filter spectrum. (Right) Low frequency filter spectrum.

tain an  $N/2$  length approximation of  $z$ , while the components of  $z$  represented over the translations of  $v$  will contain the details of  $z$ . Figure 1.12 shows the representation of  $z$  in the wavelet basis. The first half is an approximation of  $z$ , while the second contains the details needed to reconstruct  $z$ . We can revise the compression algorithm to discard the detail components of the wavelet representation of  $z$ . This will (most of the time) maintain the original signal to noise ratio of  $z$ , the first objection we raised to the Fourier compression algorithm. But by discarding only details, we are limited to 50% compression. By decomposing the  $N/2$  wavelet approximation over an  $N/2$  length wavelet basis, and iterating this process, we can continue to produce more detail components which can be discarded to compress the signal. Our new algorithm to compress  $z$  by  $p$  percent and transmit is as follows.

Let  $\ell$  be the minimal integer satisfying  $p \leq 1 - 2^{-\ell}$  and assume  $N$ , the length of  $z$ , satisfies  $2^\ell | N$ . Let  $z_{(0)} = z$ . Let  $p' = \frac{p}{1-2^{-\ell}}$ .

1. While  $1 \leq k \leq \ell$ 
  - a. Form  $z_{(k)}$  from  $z_{(k-1)}$  by replacing the length low frequency approximation of  $z_{(k-1)}$  (the first  $2^{-k}N$  components) by its representation over an  $2^{-k}N$  length first stage wavelet basis.
2. Order the detail components of  $z_{(\ell)}$  (the last  $N(1 - 2^{-\ell})^{-1}$  components) by their moduli and calculate  $a$ , their  $p'^{th}$  percentile.
3. If a detail component of  $z_{(\ell)}$  is less than  $a$  in modulus, discard it and replace it with 0.
4. Transmit the positions and values of the non-zero positions of  $z_{(\ell)}$ .
5. Reconstruct the  $N$  length approximation of  $z$  by taking the inverse wavelet transformations.

(In fact, a  $\ell^{th}$  stage wavelet basis can be calculated directly so that the calculation of  $z_{(\ell)}$  need not be performed iteratively.) Because this new algorithm retains the low frequency approximation coefficients, it retains a signal to noise ratio in the compressed signal better than the Fourier algorithm. The localization of the wavelet basis means we may begin reconstruction while only a few components have been transmitted. Figures 1.12 through 1.19 show the iterative calculation of  $z_{(\ell)}$  using Daubechies's D4 wavelet, a 90% and 95% compression, and their reconstructed approximations of  $z$ . Again the compressed signals are shown in blue over the original in red. Listen to the signal by pressing the play button and compare it with the Fourier method. While the same amount of compression is achieved with both, the wavelet compressed signal has discarded much of the background noise, while the Fourier method sounds like it has amplified it. Figure 1.20 shows a comparison of the error<sup>1</sup> between the Fourier

---

<sup>1</sup>The percent error for a compressed  $z'$  from signal  $z$  is given by  $\text{error}(z', z) = \frac{\|z' - z\|}{\|z\|}$  where  $\|\cdot\|$  is the vector space norm. See the later discussion.

and wavelet compression techniques. Observe that in this example, the Fourier method has better error values for lower levels of compression, although for compression over 80%, the wavelet method produces lower error. Compressing with a higher stage basis also appears to continue to improve the error of the compression. Bear in mind that because the wavelet compression method discards more noise information, a portion of its error is irrelevant to our purpose, and in some contexts even a desirable denoising. Then a wavelet compression approach out-performs the Fourier compression in terms of high-compression error, localization, and signal-to-noise constancy.

The iterative levels that are naturally constructed by wavelet bases lend themselves to many other applications. Because the compression separates a rough approximation and detail information, it enables algorithms which examine first rough approximations of signals then further examine the details of signals if something of interest is revealed, as might be of interest to remote sensing or image recognition applications. Many uses of wavelets are found in the literature. In the discussion that follows we will look instead towards the theory surrounding their existence and construction.

Figure 1.12: 1st Stage Wavelet Representation  $z_{(1)}$

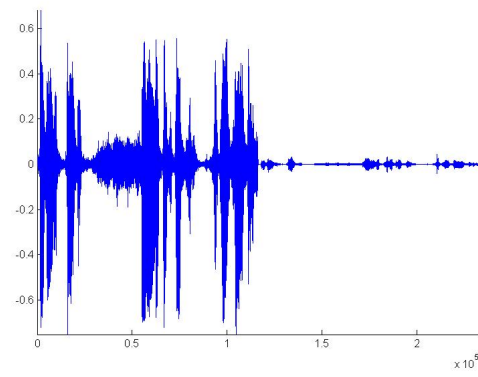


Figure 1.13: 2st Stage Wavelet Representation  $z_{(2)}$

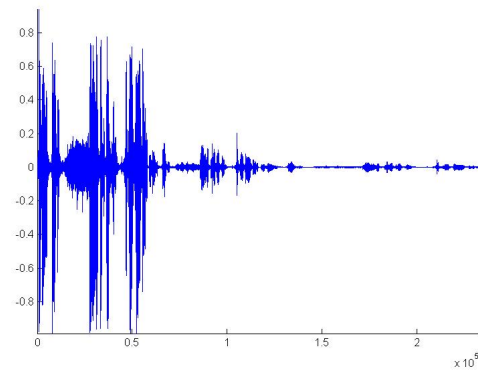




Figure 1.14: 3rd Stage Wavelet Representation  $z_{(3)}$

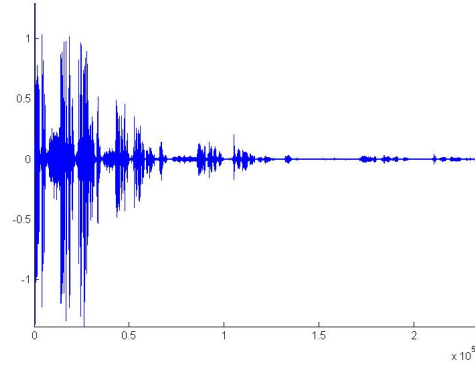


Figure 1.15: 4th Stage Wavelet Representation  $z_{(4)}$

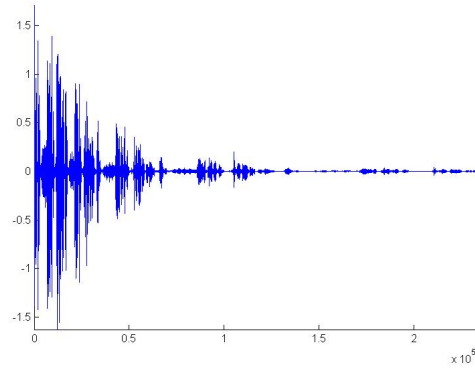


Figure 1.16: 90% Compressed 4th Stage Wavelet Representation  $z_{(4)}$

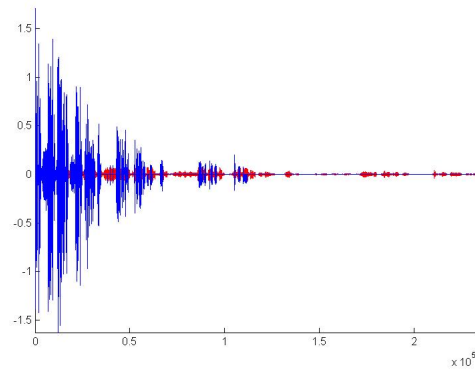
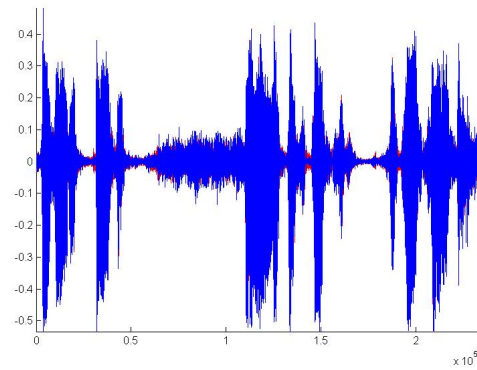


Figure 1.17: 90% Wavelet Compressed  $z$



Play

Figure 1.18: 95% Compressed 5th Stage Wavelet Representation  $z_{(5)}$

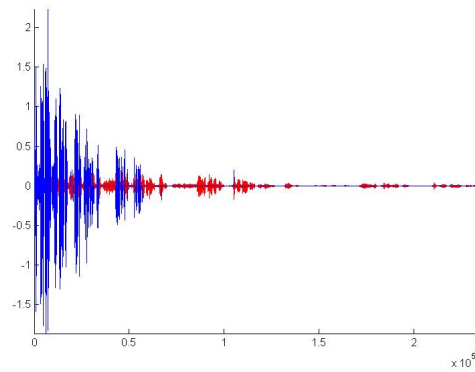
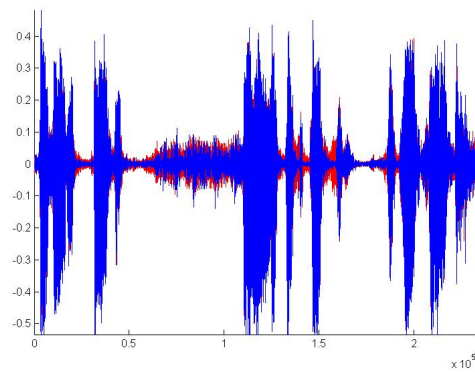
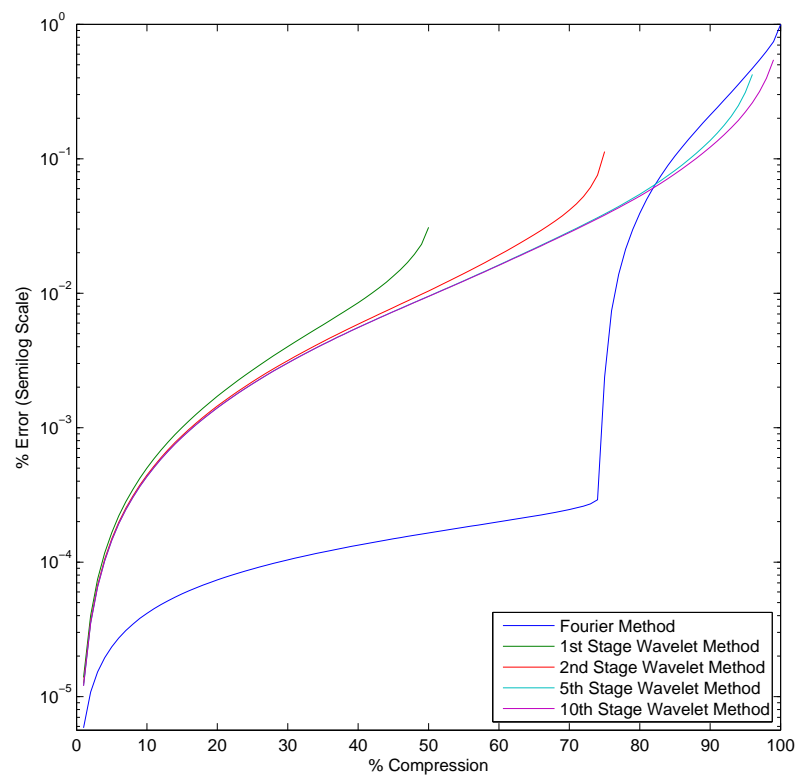


Figure 1.19: 95% Compressed  $z_{(5)}$



Play

Figure 1.20: Comparison % Error by Compression Ratio



## Chapter 2

# Discrete Wavelets

Wavelets over the spaces  $\ell^2(\mathbb{Z})$  and  $\ell^2(\mathbb{Z}_N)$  are widely used in application, and we shall build our theory of continuous wavelets off the discrete foundation. We begin by defining our notation.

For a set  $A$ , we write its power set as  $\mathcal{P}(A) = \{U \mid U \subseteq A\}$ . We will write  $\mathbb{Z}$  to refer the integers, and  $\mathbb{R}$  to refer to the field of real numbers.  $\mathbb{Z}^+$  and  $\mathbb{R}^+$  will refer to the positive integers and real numbers, respectively. We write  $\mathbb{C}$  for the field of complex numbers. For a complex number  $z = x + iy$  with  $x, y \in \mathbb{R}$  and  $i = \sqrt{-1}$  we write the complex conjugate of  $z$  as  $\bar{z} = x - iy$ .

An infinite, discrete signal is usually thought of as a vector in

$$\ell^2(\mathbb{Z}) = \left\{ v : \mathbb{Z} \rightarrow \mathbb{C} \mid \sum_{n \in \mathbb{Z}} |v(n)|^2 < \infty \right\}.$$

Most of the concepts needed for a discussion of wavelets in  $\ell^2(\mathbb{Z})$  are directly analogous to those of  $\ell^2(\mathbb{Z}_N)$ , the space of length  $N$  complex vectors, though the situation is slightly complicated by concerns of convergence. We will discuss the infinite case, and note that the finite length case is analogous, though without the necessity of limits. It is straightforward to check that  $\ell^2(\mathbb{Z})$  is a vector space under the obvious point-wise addition and scalar multiplication. We are able to equip  $\ell^2(\mathbb{Z})$  with the inner product

$$\langle v|w \rangle = \sum_{n \in \mathbb{Z}} \overline{v(n)} w(n),$$

and the norm

$$\|v\| = \langle v|v \rangle^{\frac{1}{2}}$$

which provides us with the metric  $d(v, w) = \|v - w\|$ . We will state without proof that  $\ell^2(\mathbb{Z})$  is in fact Cauchy complete with respect to this metric, making it a complete inner product space or Hilbert space (see the appendix for more information).

We distinguish one particular sequence called the *Dirac delta*  $\delta \in \ell^2(\mathbb{Z})$  where  $\delta(n) = 1$  if  $n = 0$  and  $\delta(n) = 0$  if  $n \neq 0$ . We call  $\ell^1(\mathbb{Z})$  the space defined by

$$\ell^1(\mathbb{Z}) = \left\{ v : \mathbb{Z} \rightarrow \mathbb{C} \mid \sum_{n \in \mathbb{Z}} |v(n)| < \infty \right\}$$

for which we note  $\ell^1(\mathbb{Z}) \subset \ell^2(\mathbb{Z})$ . To see this consider  $v \in \ell^1(\mathbb{Z})$ . It must be that  $\lim_{n \rightarrow -\infty} v(n) = 0$  and  $\lim_{n \rightarrow \infty} v(n) = 0$ . Then there is some  $N \in \mathbb{Z}^+$  such that  $n \geq N$  implies  $|v(n)| < 1$ . Then if  $n > N$

$$\begin{aligned} \sum_{k=-n}^n |v(k)|^2 &= \sum_{k=-N}^N |v(k)|^2 + \sum_{N < |k| \leq n} |v(k)|^2 \\ &\leq \sum_{k=-N}^N |v(k)|^2 + \sum_{N < |k| \leq n} |v(k)| \\ &\leq \sum_{k=-N}^N |v(k)|^2 + \sum_{k \in \mathbb{Z}} |v(k)| \end{aligned}$$

So we see the the sequence of partial sums  $\sum_{k=-n}^n |v(k)|^2$  is real, monotone, and bounded, so must converge to some finite number. Hence  $v \in \ell^2(\mathbb{Z})$ . The reverse inclusion is of course false. Take, for example, the sequence  $z : \mathbb{Z} \rightarrow \mathbb{C}$  with  $z(n) = (1/n)$  for  $n \neq 0$  and  $z(0) = 0$ . We know that  $\sum_{n \in \mathbb{Z}} \frac{1}{n}$  diverges so naturally  $\sum_{k \in \mathbb{Z}} |z(k)|$  also diverges, while  $\sum_{n \in \mathbb{Z}} |z(n)|^2$  converges to  $\frac{\pi^2}{6}$ .

A number of operations on  $\ell^2(\mathbb{Z})$  will be useful in studying and constructing wavelets. The *conjugate reflection* map  $\tilde{\cdot} : \ell^2(\mathbb{Z}) \rightarrow \ell^2(\mathbb{Z})$  takes  $v$  to  $\tilde{v}$  defined by

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

The *translations maps*  $R_k : \ell^2(\mathbb{Z}) \rightarrow \ell^2(\mathbb{Z})$  are defined by

$$R_k v(n) = v(n - k)$$

where  $k \in \mathbb{Z}$ . We call  $u \in \ell^2(\mathbb{Z})$  *even translate orthonormal* if the set

$$\{R_{2k} u \mid k \in \mathbb{Z}\}$$

is orthonormal. For such a vector we will call its *sister wavelet* the sequence  $v$  defined by

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

The even translate orthonormality of  $v$  is equivalent to that of  $u$ . Observe that

the sister of the sister of  $u$  is  $u$  itself and

$$\begin{aligned}
\langle v | R_{2m} v \rangle &= \sum_{k \in \mathbb{Z}} \overline{v(k)} v(k - 2m) \\
&= \sum_{k \in \mathbb{Z}} (-1)^{k-1} u(1 - k) (-1)^{k-2m-1} \overline{u(1 - k + 2m)} \\
&= \sum_{k \in \mathbb{Z}} \overline{u(1 - k + 2m)} u(1 - k) \\
&= \sum_{k \in \mathbb{Z}} \overline{u(k)} u(k - 2m) \\
&= \langle u | R_{2m} u \rangle
\end{aligned}$$

We call a basis  $B$  of  $\ell^2(\mathbb{Z})$  a *first-stage wavelet basis* if there is some  $u \in B$  such that  $u$  is even translate orthonormal with sister wavelet  $v$  and

$$B = \{R_{2k}u \mid k \in \mathbb{Z}\} \cup \{R_{2k}v \mid k \in \mathbb{Z}\}.$$

We write  $L^2[-\pi, \pi)$  for the space of square integrable functions on  $[-\pi, \pi)$ :

$$L^2[-\pi, \pi) = \left\{ f : [-\pi, \pi) \rightarrow \mathbb{C} \mid \int_{-\pi}^{\pi} |f(t)|^2 dt < \infty \right\}.$$

We define the Fourier transform on  $\ell^2(\mathbb{Z})$  as the map  $\mathcal{F} : \ell^2(\mathbb{Z}) \rightarrow L^2[-\pi, \pi)$  such that  $\mathcal{F}(z) = \hat{z}$  defined by

$$\hat{z}(\omega) = \frac{1}{\sqrt{2\pi}} \sum_{n \in \mathbb{Z}} z(n) e^{in\omega}.$$

It can be shown that the Fourier transform has inverse  $\mathcal{F}^{-1} : L^2[-\pi, \pi) \rightarrow \ell^2(\mathbb{Z})$  where  $\mathcal{F}^{-1} : f \mapsto \check{f}$  such that

$$\check{f}(n) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(\omega) e^{-in\omega} d\omega,$$

and we have  $\hat{\hat{z}} = z$  and  $\hat{\check{f}} = f$ . If we consider that  $L^2[-\pi, \pi)$  itself is a Hilbert space (see appendix) equipped with its own inner product

$$\langle f | g \rangle = \int_{-\pi}^{\pi} \overline{f(\omega)} g(\omega) d\omega,$$

then we can write  $\hat{z}(\omega) = \left\langle \frac{e^{-in\omega}}{\sqrt{2\pi}} \mid z \right\rangle$  where the inner product is in  $\ell^2(\mathbb{Z})$  and  $\check{f}(n) = \left\langle \frac{e^{in\omega}}{\sqrt{2\pi}} \mid f \right\rangle$  where the inner product is that of  $L^2[-\pi, \pi)$ . The symmetry between the inner products of these two spaces is even greater, and given by Parseval's formula:

$$\langle z | w \rangle = \langle \hat{z} | \hat{w} \rangle$$

where again the first inner product is over  $\ell^2(\mathbb{Z})$  while the second is that of  $L^2[-\pi, \pi)$ . We note a few helpful Fourier relations:

$$\sqrt{2\pi}\hat{\delta}(\omega) = \sum_{n \in \mathbb{Z}} \delta(n)e^{in\omega} = 1$$

,

$$\begin{aligned} (-1)^n \widehat{z(n)}(\omega) &= \sum_{n \in \mathbb{Z}} (-1)^n z(n) \frac{e^{in\omega}}{\sqrt{2\pi}} \\ &= \sum_{n \in \mathbb{Z}} z(n) \frac{e^{in(\omega+\pi)}}{\sqrt{2\pi}} \\ &= \hat{z}(\omega + \pi), \end{aligned}$$

$$\begin{aligned} \widehat{R_k z}(\omega) &= \sum_{n \in \mathbb{Z}} z(n-k) \frac{e^{in\omega}}{\sqrt{2\pi}} \\ &= \sum_{n \in \mathbb{Z}} z(n) \frac{e^{in\omega}}{\sqrt{2\pi}} e^{ik\omega} \\ &= \hat{z}(\omega) e^{ik\omega}. \end{aligned}$$

The operation of *convolution* is a binary operation  $\cdot * \cdot : \ell^1(\mathbb{Z}) \times \ell^1(\mathbb{Z}) \rightarrow \ell^2(\mathbb{Z})$  where

$$\begin{aligned} z * w(n) &= \sum_{m \in \mathbb{Z}} z(n-m)w(m) \\ &= \sum_{m \in \mathbb{Z}} R_n \tilde{z}(m)w(m) \\ &= \langle R_n \tilde{z} | w \rangle \end{aligned}$$

.

The convolution has the useful property that  $\widehat{z * w} = \sqrt{2\pi}\hat{z}\hat{w}$ .

**Theorem 1.** *Let  $u \in \ell^1(\mathbb{Z})$  be even translate orthonormal. Then if we define the sister wavelet  $v$  by*

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

*Then  $B = \{R_{2k}u \mid k \in \mathbb{Z}\} \cup \{R_{2k}v \mid k \in \mathbb{Z}\}$  is an orthonormal basis for  $\ell^2(\mathbb{Z})$ , and we call  $B$  a first stage wavelet basis.*

*Proof.* Observe that if  $u$  is even translate orthonormal, then  $\langle R_{2m}u | u \rangle = \delta(m)$

and we have by Parseval's formula that

$$\begin{aligned}
\langle u | R_{2m} u \rangle &= \left\langle \hat{u} \left| \widehat{R_{2m} u} \right. \right\rangle \\
&= \int_{-\pi}^{\pi} \overline{\hat{u}(\omega)} \hat{u}(\omega) e^{-i2m\omega} d\omega \\
&= \int_{-\pi}^0 |\hat{u}(\omega)|^2 e^{-i2m\omega} d\omega + \int_0^{\pi} |\hat{u}(\omega)|^2 e^{-i2m\omega} d\omega \\
&= \int_0^{\pi} |\hat{u}(\omega + \pi)|^2 e^{-i2m\omega} d\omega + \int_0^{\pi} |\hat{u}(\omega)|^2 e^{-i2m\omega} d\omega \\
&= \int_0^{\pi} (|\hat{u}(\omega)|^2 + |\hat{u}(\omega + \pi)|^2) e^{-i2m\omega} d\omega
\end{aligned}$$

Then

$$\delta(m) = \int_{-\pi}^{\pi} \frac{1}{2} \left( \left| \hat{u}\left(\frac{\omega}{2}\right) \right|^2 + \left| \hat{u}\left(\frac{\omega}{2} + \pi\right) \right|^2 \right) e^{-im\omega} d\omega.$$

By taking the Fourier transform on both sides we have for  $\omega \in [-\pi, \pi)$

$$\frac{1}{\sqrt{2\pi}} = \frac{1}{2} \left| \hat{u}\left(\frac{\omega}{2}\right) \right|^2 + \frac{1}{2} \left| \hat{u}\left(\frac{\omega}{2} + \pi\right) \right|^2,$$

or for  $\omega \in [0, \pi)$

$$\frac{2}{\sqrt{2\pi}} = |\hat{u}(\omega)|^2 + |\hat{u}(\omega + \pi)|^2. \quad (2.1)$$

The same argument applied to  $v$  yields that

$$\frac{2}{\sqrt{2\pi}} = |\hat{v}(\omega)|^2 + |\hat{v}(\omega + \pi)|^2 \quad (2.2)$$

is equivalent with  $v$  being even translate orthonormal. Consider the Fourier transform of  $v$  calculated by its definition through  $u$

$$\begin{aligned}
\hat{v}(\omega) &= \sum_{k \in \mathbb{Z}} (-1)^{k-1} \overline{u(1-k)} \frac{e^{ik\omega}}{\sqrt{2\pi}} \\
&= \sum_{k \in \mathbb{Z}} (-1)^{-k} \overline{u(k)} \frac{e^{i(1-k)\omega}}{\sqrt{2\pi}} \\
&= e^{i\omega} \sum_{k \in \mathbb{Z}} \overline{u(k)} \frac{e^{ik(\omega+\pi)}}{\sqrt{2\pi}} \\
&= e^{i\omega} \overline{\hat{u}(\omega + \pi)}.
\end{aligned}$$

Then  $\hat{v}(\omega + \pi) = -e^{i\omega} \overline{\hat{u}(\omega)}$  and hence  $v$  satisfies equation (2.2) if and only if  $u$  satisfies equation (2.1) and we also have

$$\hat{v}(\omega) \overline{\hat{u}(\omega)} + \hat{v}(\omega + \pi) \overline{\hat{u}(\omega + \pi)} = 0 \quad (2.3)$$



is equivalent to  $\langle u | R_{2m} v \rangle = 0$ , by a similar argument as above. Observe that equations 2.1, 2.2, and 2.3 are equivalent to the matrix equation:

$$\begin{bmatrix} \bar{\hat{u}}(\omega) & \bar{\hat{u}}(\omega + \pi) \\ \bar{\hat{v}}(\omega) & \bar{\hat{v}}(\omega + \pi) \end{bmatrix} \begin{bmatrix} \hat{u}(\omega) & \hat{v}(\omega) \\ \hat{u}(\omega + \pi) & \hat{v}(\omega + \pi) \end{bmatrix} = \sqrt{\frac{2}{\pi}} I_2,$$

but then these matrices commute, so

$$\begin{bmatrix} \hat{u}(\omega) & \hat{v}(\omega) \\ \hat{u}(\omega + \pi) & \hat{v}(\omega + \pi) \end{bmatrix} \begin{bmatrix} \bar{\hat{u}}(\omega) & \bar{\hat{u}}(\omega + \pi) \\ \bar{\hat{v}}(\omega) & \bar{\hat{v}}(\omega + \pi) \end{bmatrix} = \sqrt{\frac{2}{\pi}} I_2. \quad (2.4)$$

Fix some  $z \in \ell^2(\mathbb{Z})$  and let  $w \in \ell^2(\mathbb{Z})$  be given by

$$w = \sum_{k \in \mathbb{Z}} \langle R_{2k} v | z \rangle R_{2k} v + \sum_{k \in \mathbb{Z}} \langle R_{2k} u | z \rangle R_{2k} u.$$

We take the Fourier transform of  $f$ . Because the Fourier transform is a continuous and linear operator

$$\begin{aligned} \hat{w}(\omega) &= \sum_{k \in \mathbb{Z}} \langle R_{2k} v | z \rangle \widehat{R_{2k} v}(\omega) + \sum_{k \in \mathbb{Z}} \langle R_{2k} u | z \rangle \widehat{R_{2k} u}(\omega) \\ &= \sum_{k \in \mathbb{Z}} \langle R_{2k} v | z \rangle \hat{v}(\omega) e^{i2k\omega} + \sum_{k \in \mathbb{Z}} \langle R_{2k} u | z \rangle \hat{u}(\omega) e^{i2k\omega} \\ &= \hat{v}(\omega) \sum_{k \in \mathbb{Z}} \langle R_{2k} v | z \rangle e^{i2k\omega} + \hat{u}(\omega) \sum_{k \in \mathbb{Z}} \langle R_{2k} u | z \rangle e^{i2k\omega}. \end{aligned} \quad (2.5)$$

Observe that as  $\langle R_{2k} v | z \rangle = \tilde{v} * z(2k)$  we can take

$$\begin{aligned} \sum_{k \in \mathbb{Z}} \langle R_{2k} v | z \rangle e^{i2k\omega} &= \sum_{k \in \mathbb{Z}} \tilde{v} * z(2k) e^{i2k\omega} \\ &= \sum_{k \in \mathbb{Z}} \frac{1}{2} (\tilde{v} * z(k) + (-1)^k \tilde{v} * z(k)) e^{ik\omega} \\ &= \frac{\sqrt{2\pi}}{2} \sum_{k \in \mathbb{Z}} \tilde{v} * z(k) \frac{e^{ik\omega}}{\sqrt{2\pi}} + \frac{\sqrt{2\pi}}{2} \sum_{k \in \mathbb{Z}} \tilde{v} * z(k) \frac{e^{ik(\omega+\pi)}}{\sqrt{2\pi}} \\ &= \frac{\sqrt{2\pi}}{2} \widehat{\tilde{v} * z}(\omega) + \frac{\sqrt{2\pi}}{2} \widehat{\tilde{v} * z}(\omega + \pi) \\ &= \pi \bar{\tilde{v}}(\omega) \hat{z}(\omega) + \pi \bar{\tilde{v}}(\omega + \pi) \hat{z}(\omega + \pi). \end{aligned} \quad (2.6)$$

An identical derivation with  $u$  yields

$$\sum_{k \in \mathbb{Z}} \langle R_{2k} u | z \rangle e^{i2k\omega} = \pi \bar{\tilde{u}}(\omega) \hat{z}(\omega) + \pi \bar{\tilde{u}}(\omega + \pi) \hat{z}(\omega + \pi). \quad (2.7)$$

Substitute equations (2.6) and (2.7) into (2.5) we find

$$\begin{aligned}\hat{w}(\omega) &= \pi \hat{v}(\omega)(\bar{\hat{v}}(\omega)\hat{z}(\omega) + \bar{\hat{v}}(\omega + \pi)\hat{z}(\omega + \pi)) + \pi \hat{u}(\omega)(\bar{\hat{u}}(\omega)\hat{z}(\omega) + \bar{\hat{u}}(\omega + \pi)\hat{z}(\omega + \pi)) \\ &= \hat{z}(\omega)\pi(|\hat{u}(\omega)|^2 + |\hat{v}(\omega)|^2) + \hat{z}(\omega + \pi)\pi(\hat{v}(\omega)\bar{\hat{v}}(\omega + \pi) + \hat{u}(\omega)\bar{\hat{u}}(\omega + \pi))\end{aligned}$$

But equation 2.4 says that

$$\begin{aligned}\pi(|\hat{u}(\omega)|^2 + |\hat{v}(\omega)|^2) &= \sqrt{2\pi} \\ \pi(\hat{v}(\omega)\bar{\hat{v}}(\omega + \pi) + \hat{u}(\omega)\bar{\hat{u}}(\omega + \pi)) &= 0\end{aligned}$$

so we have  $\hat{w} = \hat{z}$  and by the Fourier inversion

$$z = \sum_{k \in \mathbb{Z}} \langle R_{2k}v | z \rangle R_{2k}v + \sum_{k \in \mathbb{Z}} \langle R_{2k}u | z \rangle R_{2k}u$$

so  $\ell^2(\mathbb{Z}) \subseteq \text{span } B$ , that is  $\{R_{2k}u \mid k \in \mathbb{Z}\} \cup \{R_{2k}v \mid k \in \mathbb{Z}\}$  is an orthonormal basis of  $\ell^2(\mathbb{Z})$ .  $\square$

During the proof of Theorem 1 we saw that if the even translations of  $u$  and  $v$  form a first stage wavelet basis then their Fourier transforms satisfy:

$$\begin{aligned}|\hat{u}(\omega)|^2 + |\hat{u}(\omega + \pi)|^2 &= \sqrt{2\pi}, \\ |\hat{v}(\omega)|^2 + |\hat{v}(\omega + \pi)|^2 &= \sqrt{2\pi}.\end{aligned}$$

It is convenient if we choose  $u(0) = (2\pi)^{\frac{1}{4}}$  and  $v(\pi) = (2\pi)^{\frac{1}{4}}$ . Then we will call  $u$  the low frequency filter and  $v$  the high frequency filter. It is this choice which enables the simultaneous time-frequency localization of the wavelet basis. For any  $z \in \ell^2(\mathbb{Z})$  we have

$$z = \sum_{k \in \mathbb{Z}} \langle R_{2k}v | z \rangle R_{2k}v + \sum_{k \in \mathbb{Z}} \langle R_{2k}u | z \rangle R_{2k}u.$$

The first term contains the high frequency adjustments to  $z$  and represents details of the signal, while the second term contains a low frequency approximation of the signal. Note that because any  $z$  can be written in this form, we can in particular write

$$R_m\delta = \sum_{k \in \mathbb{Z}} \langle R_{2k}v | R_m\delta \rangle R_{2k}v + \sum_{k \in \mathbb{Z}} \langle R_{2k}u | R_m\delta \rangle R_{2k}u.$$

We can reduce with  $\langle R_{2k}v | R_m\delta \rangle = \overline{v(m - 2k)}$  and  $\langle R_{2k}u | R_m\delta \rangle = \overline{u(m - 2k)}$ , so

$$\delta(n - m) = \sum_{k \in \mathbb{Z}} (\overline{v(m - 2k)}v(n - 2k) + \overline{u(m - 2k)}u(n - 2k)), \quad (2.8)$$

which is a useful equation we will utilize in proving construction of wavelets in the continuous case.

We have seen that a first stage wavelet basis can be constructed any time we can find an even translate orthonormal sequence in  $\ell^1(\mathbb{Z})$ . In the next chapter we shall see that continuous wavelets can also be constructed from such a  $\ell^1(\mathbb{Z})$  sequence.

## Chapter 3

# Continuous Wavelets

Just as in the discrete case, we are interested in forming bases of  $L^2(\mathbb{R})$  generated by simple transformations of a single function.  $L^2(\mathbb{R})$  is the vector space of square Lebesgue integrable functions

$$L^2(\mathbb{R}) = \left\{ f : \mathbb{R} \rightarrow \mathbb{C} \mid \int_{t \in \mathbb{R}} |f(t)|^2 dt < \infty \right\}$$

with inner product

$$\langle f | g \rangle = \int_{t \in \mathbb{R}} \overline{f(t)} g(t) dt$$

and norm  $\|f\| = \langle f | f \rangle^{\frac{1}{2}}$ . See the appendices on the Lebesgue measure, integration, and  $L^2(\mathbb{R})$  for more information. For  $f \in L^2(\mathbb{R})$  we define its  $j^{th}$  dilate and  $k^{th}$  translate as the  $L^2(\mathbb{R})$  function  $f_{j,k}$  defined by

$$f_{j,k}(t) = 2^{j/2} f(2^j t - k).$$

One can easily check that  $L^2(\mathbb{R})$  is closed under these transformations.

$$\begin{aligned} \|f_{j,k}\| &= \int_{t \in \mathbb{R}} \overline{2^{j/2} f(2^j t - k)} 2^{j/2} f(2^j t - k) dt \\ &= \int_{t \in \mathbb{R}} 2^j |f(2^j t - k)|^2 dt \\ &= \int_{t \in \mathbb{R}} |f(t)|^2 dt \\ &= \|f\|. \end{aligned}$$

We will call  $\psi \in L^2(\mathbb{R})$  a *wavelet* if the set of all its translates and dilates  $\{\psi_{j,k} \mid j, k \in \mathbb{Z}\}$  forms an orthonormal basis of  $L^2(\mathbb{R})$ .

**Definition 2.** A multi-resolution analysis  $V$  is a sequence of subspaces of  $L^2(\mathbb{R})$   $V : \mathbb{Z} \rightarrow \mathcal{P}(L^2(\mathbb{R}))$  satisfying the following properties:

1. Space:  $V_j$  is a subspace of  $L^2(\mathbb{R})$ .
2. Non-decreasing: If  $j < k$  then  $V_j \subseteq V_k$ .
3. Density: The set  $\bigcup_{j \in \mathbb{Z}} V_j$  is dense in  $L^2(\mathbb{R})$ .
4. Trivial Intersection:  $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$ .
5. Dilation:  $f \in V$  if and only if  $f_{j,0} \in V_j$ .
6. Scaling: There is some  $\phi \in V_0$  called the scaling function such that  $\{\phi_{0,k} \mid k \in \mathbb{Z}\}$  is an orthonormal basis for  $V_0$ .

The multi-resolution analysis has a number of useful properties, and we will eventually use them to construct wavelets over  $L^2(\mathbb{R})$ . Observe first that the dilation and scaling properties of a multi-resolution analysis  $V$  will guarantee that the set of translations of the  $j^{\text{th}}$  dilate of  $\phi$ , the scaling function of  $V$ , is an orthonormal basis for  $V_j$ . By the dilation property of  $V$  we have that  $\phi_{0,k} \in V_0$  implies  $\phi_{j,k} \in V_j$ . Naturally

$$\begin{aligned} \langle \phi_{j,k'} | \phi_{j,k} \rangle &= \int_{\mathbb{R}} \overline{2^{j/2} \phi(2^j t - k')} 2^{j/2} \phi(2^j t - k) dt \\ &= \int_{\mathbb{R}} \overline{\phi(t)} \phi(t - (k - k')) dt \\ &= \langle \phi | \phi_{0,k-k'} \rangle, \end{aligned}$$

so it follows that  $\{\phi_{j,k} \mid k \in \mathbb{Z}\}$  is orthonormal. And if  $f \in V_j$  then the dilation property of  $V$  asserts that  $f = g_{j,0}$  for some  $g \in V_0$ . But because the translations of  $\phi$  form an orthonormal basis of  $V_0$  we have:

$$g = \sum_{k \in \mathbb{Z}} \langle \phi_{0,k} | g \rangle \phi_{0,k}$$

so that by taking the dilate of  $g$  we obtain for almost every  $t \in \mathbb{R}$ :

$$\begin{aligned} f(t) &= 2^{j/2} g(2^j t) \\ &= 2^{j/2} \sum_{k \in \mathbb{Z}} \langle \phi_{0,k} | g \rangle \phi_{0,k}(2^j t) \\ &= \sum_{k \in \mathbb{Z}} \langle \phi_{0,k} | g \rangle \phi_{j,2^j k}(t). \end{aligned}$$

Hence  $f$  is in the span of  $\{\phi_{j,k} \mid k \in \mathbb{Z}\}$ , which must constitute an orthonormal basis of  $V_j$ . Observe that as  $V$  is non-decreasing we have  $\phi \in V_0 \subseteq V_1$  and so we have

$$\phi = \sum_{k \in \mathbb{Z}} \langle \phi_{1,k} | \phi \rangle \phi_{1,k}.$$

We will define  $u \in \ell^2(\mathbb{Z})$ , called the scaling sequence of  $V$ , satisfying  $u(k) = \langle \phi_{1,k} | \phi \rangle$ . That is

$$\phi = \sum_{k \in \mathbb{Z}} u(k) \phi_{1,k},$$

and we have

$$\begin{aligned}
\phi_{0,m}(t) &= \sum_{k \in \mathbb{Z}} u(k) \phi_{1,k}(t - m) \\
&= \sum_{k \in \mathbb{Z}} u(k) 2^{1/2} \phi(2t - 2m - k) \\
&= \sum_{k \in \mathbb{Z}} u(k - 2m) 2^{1/2} \phi(2t - k) \\
&= \sum_{k \in \mathbb{Z}} R_{2m} u(k) \phi_{1,k}(t).
\end{aligned}$$

One can see that  $u$  is even translate orthonormal in  $\ell^2(\mathbb{Z})$ . Indeed, observe that

$$\langle \phi | \phi_{0,m} \rangle = \left\langle \sum_{k \in \mathbb{Z}} u(k) \phi_{1,k} \left| \sum_{k' \in \mathbb{Z}} R_{2m} u(k') \phi_{1,k'} \right. \right\rangle.$$

If we utilize the continuity and linearity of the inner product

$$\begin{aligned}
&= \sum_{k \in \mathbb{Z}} \sum_{k' \in \mathbb{Z}} \overline{u(k)} R_{2m} u(k') \langle \phi_{1,k} | \phi_{1,k'} \rangle \\
&= \sum_{k \in \mathbb{Z}} \sum_{k' \in \mathbb{Z}} \overline{u(k)} R_{2m} u(k') \delta(k - k') \\
&= \sum_{k \in \mathbb{Z}} \overline{u(k)} R_{2m} u(k) \\
&= \langle u | R_{2m} u \rangle,
\end{aligned}$$

where this last inner product is over  $\ell^2(\mathbb{Z})$ . So because the translations of  $\phi$  are orthonormal, it follows that the even translations of  $u$  are orthonormal in  $\ell^2(\mathbb{Z})$ . Remember that according to Theorem 1 we have that if  $u \in \ell^1(\mathbb{Z})$  then its sister wavelet  $v$  defined by

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

is in  $\ell^1(\mathbb{Z})$  and the even translations of  $u$  and  $v$  form a first stage wavelet basis in  $\ell^2(\mathbb{Z})$ . The first stage wavelet basis generated by  $u$  and  $v$  will allow us to construct a wavelet basis in  $L^2(\mathbb{R})$  itself.

**Theorem 3** (Mallat). *Let  $V : \mathbb{Z} \rightarrow \mathcal{P}(L^2(\mathbb{R}))$  be a multi-resolution analysis with scaling function  $\phi$  and scaling sequence  $u$  with sister wavelet of  $v$ . Define*

$$\psi = \sum_{k \in \mathbb{Z}} v(k) \phi_{1,k}.$$

*Then  $\psi$  is a wavelet for  $L^2(\mathbb{R})$ .*

*Proof.*  $\psi$  is a wavelet for  $L^2(\mathbb{R})$  if  $\{\psi_{j,k} \mid j, k \in \mathbb{Z}\}$  is an orthonormal basis for  $L^2(\mathbb{R})$ . We first show that for a fixed  $j$ ,  $W_j = \{\psi_{j,k} \mid k \in \mathbb{Z}\}$  is orthonormal. Observe that

$$\begin{aligned}\langle \psi_{j,k'} \mid \psi_{j,k} \rangle &= \int_{\mathbb{R}} \overline{2^{j/2} \psi(2^j t - k')} 2^{j/2} \psi(2^j t - k) dt \\ &= \int_{\mathbb{R}} \overline{\psi(t)} 2^{j/2} \psi(t - (k - k')) dt \\ &= \langle \psi \mid \psi_{0,k-k'} \rangle.\end{aligned}$$

It will then suffice to show that  $\langle \psi \mid \psi_{0,m} \rangle = \delta(m)$ . If we expand  $\psi$  over the same  $\phi_{1,k}$  which define  $\psi$ , then we can write  $\psi_{j,m}$  in terms of the basis of  $V_{j+1}$ .

$$\begin{aligned}\psi_{j,m}(x) &= 2^{j/2} \psi(2^j t - m) \\ &= 2^{j/2} \sum_{k \in \mathbb{Z}} v(k) \phi_{1,k}(2^j t - m) \\ &= 2^{j/2} \sum_{k \in \mathbb{Z}} v(k) 2^{1/2} \phi(2^{j+1} t - 2m - k) \\ &= \sum_{k \in \mathbb{Z}} v(k - 2m) 2^{(j+1)/2} \phi(2^{j+1} t - k) \\ &= \sum_{k \in \mathbb{Z}} R_{2m} v(k) \phi_{j+1,k}(t).\end{aligned}$$

The continuity of the inner product allows us to take

$$\begin{aligned}\langle \psi \mid \psi_{0,m} \rangle &= \left\langle \sum_{k \in \mathbb{Z}} v(k) \phi_{1,k} \mid \sum_{k' \in \mathbb{Z}} R_{2m} v(k') \phi_{1,k'} \right\rangle \\ &= \sum_{k \in \mathbb{Z}} \sum_{k' \in \mathbb{Z}} \overline{v(k)} R_{2m} v(k') \langle \phi_{1,k} \mid \phi_{1,k'} \rangle.\end{aligned}$$

but we know that  $\langle \phi_{1,k} \mid \phi_{1,k'} \rangle = \langle \phi_{0,k} \mid \phi_{0,k'} \rangle = \delta(k - k')$  follows from the definition of the scaling function. Then

$$\begin{aligned}\langle \psi \mid \psi_{0,m} \rangle &= \sum_{k \in \mathbb{Z}} \sum_{k' \in \mathbb{Z}} \overline{v(k)} R_{2m} v(k') \delta(k - k') \\ &= \sum_{k \in \mathbb{Z}} \overline{v(k)} R_{2m} v(k) \\ &= \langle v \mid R_{2m} v \rangle,\end{aligned}$$

where the final inner product is in  $\ell^2(\mathbb{Z})$ . But we saw in Theorem 1 that  $v$  is even translate orthonormal if and only if  $u$  is even translate orthonormal. Hence the orthonormality of  $\{\psi_{j,k} \mid k \in \mathbb{Z}\}$  reduces to the orthonormality of the even translates of  $u$ , which in turn follows from the orthonormality of  $\{\phi_{0,k} \mid k \in \mathbb{Z}\}$ .

We may then define  $W_j = \text{span}\{\psi_{j,k} \mid k \in \mathbb{Z}\}$  which is a subspace of  $L^2(\mathbb{R})$ . We see that for fixed  $j \in \mathbb{Z}$  by following the same style of argument as above

$$\langle \phi_{j,k} \mid \psi_{j,k'} \rangle = \langle R_{2k} u \mid R_{2k'} v \rangle = 0.$$

Then  $W_j \perp V_j$ . We will next aim to show that  $V_j \oplus W_j = V_{j+1}$ .

Because  $V$  is a multi-resolution analysis  $V_j \subseteq V_{j+1}$  and by the expansion  $\psi_{j,m} = \sum_{k \in \mathbb{Z}} R_{2m} v(k) \phi_{j+1,k}$  we have that  $\psi_{j,k} \in V_{j+1}$  for all  $k \in \mathbb{Z}$ . Then because  $\{\psi_{j,k} | k \in \mathbb{Z}\} \subseteq V_{j+1}$  we have that  $W_j = \text{span}\{\psi_{j,k} | k \in \mathbb{Z}\} \subseteq V_{j+1}$ , so  $V_j \oplus W_j \subseteq V_{j+1}$ . Let  $g \in V_j \oplus W_j$  be given by

$$\begin{aligned} g &= \sum_{m \in \mathbb{Z}} R_{2m} \overline{u(k)} \phi_{j,m} + \sum_{m \in \mathbb{Z}} R_{2m} \overline{v(k)} \psi_{j,m} \\ &= \sum_{m \in \mathbb{Z}} R_{2m} \overline{u(k)} \sum_{n \in \mathbb{Z}} R_{2m} u(n) \phi_{j+1,n} + \sum_{m \in \mathbb{Z}} R_{2m} \overline{v(k)} \sum_{n \in \mathbb{Z}} R_{2m} v(n) \phi_{j+1,n} \\ &= \sum_{n \in \mathbb{Z}} \left( \sum_{m \in \mathbb{Z}} (\overline{u(k-2m)} u(n-2m) + \overline{v(k-2m)} v(n-2m)) \right) \phi_{j+1,n} \end{aligned}$$

The middle summation is a Dirac delta by equation 2.8. We are left with

$$\begin{aligned} g &= \sum_{n \in \mathbb{Z}} \delta(k-n) \phi_{j+1,n} \\ &= \phi_{j+1,k}. \end{aligned}$$

Then as the defining basis of  $V_{j+1}$  is the sum of terms in  $V_j$  and  $W_j$ , we have that  $V_{j+1} = V_j \oplus W_j$ . For any  $m < j$  we have  $W_m \subseteq V_{m+1} \subseteq \dots \subseteq V_j$  so we have  $\psi_{j,k} \in V_{j+1}$ , and because  $W_j \perp V_j$  we see that  $\psi_{j,k'} \perp \psi_{m,k}$  for all  $k, k' \in \mathbb{Z}$ . It follows that  $\{\psi_{j,k} | j, k \in \mathbb{Z}\}$  is orthonormal in  $L^2(\mathbb{R})$ .

All that remains is to show that  $\{\psi_{j,k} | j, k \in \mathbb{Z}\}$  is complete. Assume that there is some  $f \in L^2(\mathbb{R})$  such that  $\langle f | \psi_{j,k} \rangle = 0$  for every  $j, k \in \mathbb{Z}$ . Let  $P_j(f)$  be the projection of  $f$  onto  $V_j$  defined by

$$P_j(f) = \sum_{k \in \mathbb{Z}} \langle \phi_{j,k} | f \rangle \phi_{j,k}.$$

Observe that  $f - P_j(f) \perp V_j$  so for every  $\ell < j$ :  $f - P_j(f) \perp W_\ell$  and as by assumption  $f \perp W_\ell$  we see  $P_j(f) \perp W_\ell$ . Because  $V_j = V_{j-1} \oplus W_{j-1}$  and  $P_j(f) \perp W_{j-1}$  we see that  $P_j(f) \in V_{j-1}$  and by induction  $g \in V_\ell$  for every  $\ell < j$ . As  $V$  is non-decreasing  $P_j(f) \in \bigcap_{j \in \mathbb{Z}} V_j$ . Then the trivial intersection property of  $V$  forces  $P_j(f) = 0$ . Then

$$\|f\| = \|f - P_j(f)\| = \inf \{\|f - v\| \mid v \in V_j\}$$

for every  $j \in \mathbb{Z}$ . By the density property of  $V$  we have that there is some sequence  $g : \mathbb{Z}^+ \rightarrow \bigcup_{j \in \mathbb{Z}} V_j$  such that  $g_n$  converges to  $f$  in the  $L^2(\mathbb{R})$  metric. But of course as  $g_n \in V_j$  for some  $j \in \mathbb{Z}$  and  $\|f\| = \inf \{\|f - v\| \mid v \in V_j\}$  we see that

$$\|f\| \leq \|f - g_n\|$$

for any  $n \in \mathbb{Z}$ . Then  $f = 0$ , and it must be that  $\{\psi_{j,k} | j, k \in \mathbb{Z}\}$  is a basis of  $L^2(\mathbb{R})$ .  $\square$

In contrast with the discrete case, wavelets in  $L^2(\mathbb{R})$  have a bi-infinite dimension of dilations, and there is no clear base level off which to form an approximation. But if we distinguish some  $\ell \in \mathbb{Z}$  then in fact the set  $B'$  given by

$$B' = \{\phi_{\ell,k} \mid k \in \mathbb{Z}\} \cup \{\psi_{j,k} \mid k \in \mathbb{Z}, j \geq \ell\}$$

is an orthonormal basis for  $L^2(\mathbb{R})$ . That  $B'$  is orthonormal should be clear from the above discussion of the orthonormal translates of  $\phi$  and the orthonormal translates and dilates of  $\psi$ . Suppose there is some  $g \in L^2(\mathbb{R})$  such that  $g \perp B'$ . By Theorem 3 we can write  $g$  as

$$g = \sum_{j < \ell} \sum_{k \in \mathbb{Z}} \langle \psi_{j,k} | f \rangle \psi_{j,k},$$

and so  $g \in \bigoplus_{j < \ell} W_j \subseteq V_\ell$ , where  $W$  and  $V$  are as defined in the proof of Theorem 3. As  $\{\phi_{\ell,k} \mid k \in \mathbb{Z}\}$  is an orthonormal basis for  $V_\ell$  we see that  $g = 0$ . Then any  $f \in L^2(\mathbb{R})$  can be written as

$$f = \sum_{k \in \mathbb{Z}} \langle \phi_{\ell,k} | f \rangle \phi_{\ell,k} + \sum_{j \geq \ell} \sum_{k \in \mathbb{Z}} \langle \psi_{j,k} | f \rangle \psi_{j,k}$$

where the first term is a  $\ell$  level approximation of  $f$  and the second term contains the finer details.

Theorem 3 allows us to construct a wavelet from a multi-resolution analysis, but as the multi-resolution analysis presupposes many properties, one might hazard that they are difficult to construct. In fact however, many translate orthonormal  $u \in \ell^1(\mathbb{Z})$ , provided they satisfy some additional conditions, can be used to construct a multi-resolution analysis by defining  $\phi$  via its Fourier transform

$$\hat{\phi}(\omega) = \prod_{j \in \mathbb{Z}^+} \frac{1}{\sqrt{2}} \sum_{k \in \mathbb{Z}} u(k) \frac{e^{-ik\omega/2^j}}{\sqrt{2\pi}}.$$

Though we refer the reader to [2] for proof,  $\phi$  is the scaling sequence of a multi-resolution analysis  $V$ , defined by  $V_j = \text{span}\{\phi_{j,k} \mid k \in \mathbb{Z}\}$ .

Thus a wide variety of bases for spaces housing signals can be found by the much simpler problem of identifying even translate orthonormal sequences in  $\ell^1(\mathbb{Z})$ . Such control allows some tailoring of a wavelet basis to the needs of an application.



## Chapter 4

# Appendix

### 4.1 Lebesgue Measure and Integration

For  $A \subseteq \mathbb{R}$  we define

$$\lambda^*(A) = \inf \left\{ \sum_{k=0}^{\infty} (b_k - a_k) \mid A \subseteq \bigcup_{k=0}^{\infty} (a_k, b_k) \right\}$$

as the *outer measure* of  $A$ . We say that  $A$  is *measurable* if

$$\forall B \subseteq \mathbb{R} : \lambda^*(B) = \lambda^*(B \cap A) + \lambda^*(B \cap A^c) \quad (4.1)$$

where  $A^c = \mathbb{R} - A$  is the complement of  $A$ . We will write  $\mathbb{M}$  for the set of measurable real sets and define the *Lebesgue measure*  $\lambda : \mathbb{M} \rightarrow \mathbb{R}_0^+ \cup \{\infty\}$  as the restriction of  $\lambda^*$  to  $\mathbb{M}$ .

A  $\sigma$ -*algebra* is a non-empty set of sets closed under complementation, countable unions, and countable intersections. We observe that  $\mathbb{M}$  is a  $\sigma$ -algebra.

**Theorem 4.** *Let  $A : \mathbb{Z}^+ \rightarrow \mathbb{M}$  be a sequence of sets in  $\mathbb{M}$ . Then:*

(a)  $A_0^c \in \mathbb{M}$

(b)  $\bigcap_{n \in \mathbb{Z}^+} A_n \in \mathbb{M}$

(c)  $\bigcup_{n \in \mathbb{Z}^+} A_n \in \mathbb{M}$

*Proof.* (a) If  $A_0 \in \mathbb{M}$  then it follows immediately from (4.1) that  $A_0^c \in \mathbb{M}$ .

(b) We first demonstrate *sub-additivity* of  $\lambda^*$ . Let  $A : \mathbb{Z}^+ \rightarrow \mathbb{M}$  be a sequence of sets in  $\mathbb{M}$ . For any  $n \in \mathbb{Z}$  we have

$$\lambda^*(A_n) = \inf \left\{ \sum_{m \in \mathbb{Z}^+} (b_m - a_m) \mid A_n \subseteq \bigcup_{m \in \mathbb{Z}} (a_m, b_m) \right\}.$$

Observe that for any  $\varepsilon > 0$  we can find some collection of open intervals  $\{(a_{m_n}, b_{m_n}) \mid m_n \in \mathbb{Z}\}$  covering  $A_n$  such that

$$\lambda^*(A_n) + \frac{\varepsilon}{2^n} > \sum_{m \in \mathbb{Z}^+} (b_{m_n} - a_{m_n}).$$

Then by taking a sum over  $n$  we have

$$\sum_{n \in \mathbb{Z}^+} \lambda^*(A_n) + \varepsilon > \sum_{n \in \mathbb{Z}^+} \sum_{m \in \mathbb{Z}^+} (b_{m_n} - a_{m_n}).$$

But because  $\bigcup_{n \in \mathbb{Z}^+} \bigcup_{m_n \in \mathbb{Z}^+} (a_{m_n}, b_{m_n})$  covers  $\bigcup_{n \in \mathbb{Z}^+} A_n$  we have

$$\begin{aligned} \sum_{n \in \mathbb{Z}^+} \lambda^*(A_n) + \varepsilon &> \inf \left\{ \sum_{m \in \mathbb{Z}^+} (b_m - a_m) \mid \bigcup_{n \in \mathbb{Z}^+} A_n \subseteq \bigcup_{m \in \mathbb{Z}} (a_m, b_m) \right\} \\ &= \lambda^* \left( \bigcup_{n \in \mathbb{Z}^+} A_n \right) \end{aligned}$$

But as this  $\varepsilon > 0$  is arbitrary we have

$$\lambda^* \left( \bigcup_{n \in \mathbb{Z}^+} A_n \right) \leq \sum_{n \in \mathbb{Z}^+} \lambda^*(A_n). \quad (4.2)$$

This is referred to as the sub-additivity of  $\lambda^*$ . Note then by the sub-additivity, if  $A \in \mathbb{M}$  for any  $B \subset \mathbb{R}$

$$\lambda^*(B) = \lambda^*((B \cap A) \cup (B \cap A^c)) \quad (4.3)$$

$$\leq \lambda^*(B \cap A) + \lambda^*(B \cap A^c) \quad (4.4)$$

Note also that it follows immediately from the definition of  $\lambda^*$  is monotone, that is if  $A_0 \subset A_1$  then  $\lambda^*(A_0) \leq \lambda^*(A_1)$ . Suppose that  $A_i, A_j \in \mathbb{M}$  then for any  $B \subset \mathbb{R}$  we have by applying (4.1):

$$\lambda^*(B \cap A_i \cap A_j) = \lambda^*(B) - \lambda^*(B \cap A_i^c) - \lambda^*(B \cap A_i \cap A_j^c)$$

and by set considerations we see that

$$\begin{aligned} \lambda^*(B \cap (A_i \cap A_j)^c) &= \lambda^*((B \cap A_i^c) \cup (B \cap A_j^c)) \\ &\leq \lambda^*(B \cap A_i^c) + \lambda^*(B \cap A_j^c) \end{aligned}$$

adding these results we obtain

$$\lambda^*(B \cap A_i \cap A_j) + \lambda^*(B \cap (A_i \cap A_j)^c) \leq \lambda^*(B) + \lambda^*(B \cap A_j^c) - \lambda^*(B \cap A_i \cap A_j^c).$$

As  $B \cap A_i \cap A_j^c \subseteq B \cap A_j^c$  we have

$$\lambda^*(B \cap A_i \cap A_j) + \lambda^*(B \cap (A_i \cap A_j)^c) \leq \lambda^*(B),$$

which with result (4.4), we have that  $A_i \cap A_j$  meets condition (4.1) and so  $A_i \cap A_j \in \mathbb{M}$ . It follows by induction that  $\mathbb{M}$  is closed under finite intersection. It remains to be shown that  $\mathbb{M}$  is closed under countable intersection. Suppose now that  $A : \mathbb{Z}^+ \rightarrow \mathbb{M}$  is a sequence of sets in  $\mathbb{M}$ . Let  $S_0 = \mathbb{R}$ ,  $S_n = \bigcap_{m=0}^n A_m$ , and  $S = \bigcap_{m \in \mathbb{Z}^+} A_m$ . We have that each  $S_n$  is measurable by the above argument, so for any  $B \subseteq \mathbb{R}$

$$\lambda^*(B) = \lambda^*(B \cap S_n) + \lambda^*(B \cap S_n^c),$$

but as  $S_n$  is a decreasing sequence and as  $\lambda^*$  is monotone we see  $\lambda^*(B \cap S_n) \geq \lambda^*(B \cap S)$  so it will suffice to show that  $\lambda^*(B \cap S^c) \leq \lim_{n \rightarrow \infty} \lambda^*(B \cap S_n^c)$ . Observe that we can write  $B \cap S^c$  as the union

$$B \cap S^c = \bigcup_{n=0}^{\infty} B \cap (S_{n+1}^c - S_n^c)$$

so utilizing the sub-additivity we have

$$\begin{aligned} \lambda^*(B \cap S^c) &\leq \sum_{n=0}^{\infty} \lambda^*(B \cap (S_{n+1}^c - S_n^c)) \\ &= \sum_{n=0}^{\infty} \lambda^*(B \cap S_{n+1}^c) - \lambda^*(B \cap S_{n+1}^c \cap S_n^c) \end{aligned}$$

by the measurability of  $S_n$ . But as  $S_n^c \subseteq S_{n+1}^c$

$$\begin{aligned} \lambda^*(B \cap S^c) &\leq \sum_{n=0}^{\infty} \lambda^*(B \cap S_{n+1}^c) - \lambda^*(B \cap S_n^c) \\ &= \lim_{n \rightarrow \infty} \lambda^*(B \cap S_{n+1}^c) - \lambda^*(B \cap S_0^c), \end{aligned}$$

but  $S_0^c = \emptyset$  which, as is easy to verify, has  $\lambda^*(\emptyset) = 0$ . It follows that

$$\lambda^*(B) \geq \lambda^*(B \cap S) + \lambda^*(B \cap S^c)$$

for any  $B \subseteq \mathbb{R}$ , so with equation 4.4 we find that  $S \in \mathbb{M}$ .

- (c) If  $A : \mathbb{Z}^+ \rightarrow \mathcal{P}(\mathbb{M})$  is a sequence in  $\mathbb{M}$  then by DeMorgan's law  $\bigcup_{n=0}^{\infty} A_n = (\bigcap_{n=0}^{\infty} A_n^c)^c \in \mathbb{M}$

□

We call a set *0 measure* if it is measurable and has Lebesgue measure of 0. The Lebesgue measure is complete in the sense that every subset of a 0 measure set is also measurable with 0 measure. To see this let  $Z \in \mathbb{M}$  has  $\lambda(Z) = 0$  and  $Y \subseteq Z$ . For any  $B \subseteq \mathbb{R}$  we have  $B \cap Y \subseteq Z$ , so  $\lambda^*(B \cap Y) \leq \lambda^*(Z) = 0$ . Then

$$\lambda^*(B) \geq \lambda^*(B \cap Y^c) = \lambda^*(B \cap Y) + \lambda^*(B \cap Y^c)$$

which with the subadditivity of  $\lambda^*$  is sufficient to show that  $Y \in \mathbb{M}$ . And  $\lambda(Y) \leq \lambda(Z) = 0$ .

We say that  $f : \mathbb{R} \rightarrow \mathbb{R}$  is *measurable* if the set  $\{x \in \mathbb{R} \mid f(x) > a\}$  is measurable for every  $a \in \mathbb{R}$ . We say  $f$  is *simple* if  $f$  is measurable and its range is finite. For a measurable  $f$  such that  $f \geq 0$  define the *Lebesgue integral* of  $f$  on  $A \subseteq \mathbb{R}$  as

$$\int_A f(x)dx = \sup \left\{ \sum_{n=0}^k c_n \lambda(A \cap s^{-1}(c_n)) \mid s \text{ simple } 0 \leq s \leq f \right\}$$

where  $\text{range } s = \{c_1, \dots, c_k\}$ . We say a non-negative measurable function is integrable on  $A$  if its integral is finite. If  $f$  takes negative values then let  $f^+ = \max(f, 0)$  and  $f^- = -\min(f, 0)$ , and define the integral as

$$\int_A f(x)dx = \int_A f^+(x)dx - \int_A f^-(x)dx,$$

provided both  $f^+$  and  $f^-$  are integrable, in which case we call  $f$  integrable. For  $f : \mathbb{R} \rightarrow \mathbb{C}$  we say  $f$  is measurable when its real and imaginary parts ( $\Re f$  and  $\Im f$ , respectively) are measurable, and define

$$\int_A f(x)dx = \int_A \Re f(x)dx + i \int_A \Im f(x)dx$$

provided  $\Re f$  and  $\Im f(x)$  are integrable, in which case we call  $f$  integrable. For measurable  $f : A \rightarrow \mathbb{C}$  if  $\int_A |f(x)|^p dx$  is finite we say  $f \in L^p(A)$ . But as the integral remains unchanged if we alter the function value on any set of zero measure, we will place an equivalence relation on functions in  $L^p(A)$ . We say property  $P$  holds *almost everywhere* on  $A$  (or just *a.e.*  $A$ ) when  $P$  holds for all  $x \in A$  except possibly on some measure zero set. For functions  $f, g \in L^p(A)$  we consider  $f \sim g$  if  $f(x) = g(x)$  *a.e.*  $A$  and will frequently fail to differentiate  $f \sim g$  and  $f = g$ , considering  $L^p(A)$  to be modulo the equivalence relation  $\sim$ .

If  $f$  is measurable then  $\{x \in \mathbb{R} \mid f(x) > a\}$  is measurable for every  $a \in \mathbb{R}$ . Then  $\{x \in \mathbb{R} \mid f(x) \leq a\}$  and  $\{x \in \mathbb{R} \mid f(x) < a\}$  are also measurable. Then we have  $\{x \in \mathbb{R} \mid f(x) > a\} \cap \{x \in \mathbb{R} \mid f(x) < -a\} = \{x \in \mathbb{R} \mid |f(x)| > a\}$ . So  $|f|$  is also measurable. It follows immediately that  $|f|$  is integrable whenever  $f$  is, as

$$\int_A |f(x)|dx = \int_A f^+(x)dx + \int_A f^-(x)dx,$$

Continuous functions are measurable. Continuous functions are integrable over compact sets. The sums and products of integrable functions are integrable.

## 4.2 Hilbert Spaces

A *Hilbert space*  $H$  is a vector space over the scalar field  $\mathbb{C}$ , equipped with a conjugate symmetric, linear, and positive-definite inner product  $\langle \cdot | \cdot \rangle : H \times H \rightarrow$

$\mathbb{C}$  such that  $\|\cdot\| : H \rightarrow \mathbb{R}_0^+$  with  $\|v\| = \langle v|v \rangle^{\frac{1}{2}}$  is a norm, and that this norm defines a metric  $d(v, w) = \|v - w\|$  under which  $H$  is Cauchy complete. We call a set  $A \subseteq H$  *orthonormal* if we have that for any pair of vectors  $v, w \in A$

$$\langle v|w \rangle = \begin{cases} 1 & \text{if } v = w \\ 0 & \text{if } v \neq w. \end{cases}$$

We say that  $v, w \in H$  are orthonormal and write  $v \perp w$  whenever  $\langle v|w \rangle = 0$ . Observe that  $\|v + w\|^2 = \|v\|^2 + \|w\|^2$  if and only if  $v \perp w$ .

The span of a set  $A$  is the closure in  $H$  under the metric topology of all finite linear combinations of elements of  $A$ . That is

$$\text{span } A = \text{clos}_d \left\{ \sum_{n=0}^N \alpha_n a_n \mid N \in \mathbb{Z}^+, \alpha : \mathbb{Z}^+ \rightarrow \mathbb{C}, a : \mathbb{Z}^+ \rightarrow A \right\}.$$

We say that a set  $A \subset H$  is linearly independent if for every  $a \in A$ ,  $a \notin \text{span}(A - \{a\})$ . A linearly independent set which spans  $H$  is called a *basis*. An orthonormal basis is an especially useful mathematical tool. We will state a number of useful facts of countable orthonormal bases without proof or asserting their existence in general. Suppose that  $A = \{a_n \mid a : J_A \subseteq \mathbb{Z} \rightarrow H\}$  is an orthonormal basis of  $H$  and  $B = \{b_n \mid b : J_B \subseteq \mathbb{Z} \rightarrow H\}$  is an orthonormal set then

1. If  $z \in \ell^2(\mathbb{Z})$  then  $\sum_{n \in J_B} z(n)b_n \in \text{span } B$ .
2. If  $v \in H$  has  $\langle a_n|v \rangle = 0$  for all  $n \in J_A$  then  $v = 0$ .
3. If  $v \in H$  and  $z : \mathbb{Z} \rightarrow \mathbb{C}$  has  $z(n) = \langle a_n|v \rangle$  if  $n \in J_A$  and  $z(n) = 0$  otherwise, then  $z \in \ell^2(\mathbb{Z})$ .
4. If  $v \in H$  then

$$v = \sum_{n \in J_A} \langle a_n|v \rangle a_n.$$

One important result of a general Hilbert space is the Cauchy-Schwarz inequality, which asserts that  $|\langle v|w \rangle| \leq \|v\| \|w\|$ . To see this observe that  $f : \mathbb{C} \rightarrow \mathbb{R}_0^+$  with  $f(t) = \|v - tw\|^2 \geq 0$ . Then  $f(t) = \|v\|^2 - t\langle v|w \rangle - \bar{t}\langle w|v \rangle + |t|^2\|w\|^2$ . Then

$$f\left(\frac{\langle w|v \rangle}{\|w\|^2}\right) = \|v\|^2 - \frac{|\langle w|v \rangle|^2}{\|w\|^2}$$

and the result follows.

### 4.3 $L^2(\mathbb{R})$ as a Hilbert Space

In fact

$$L^2(A) = \left\{ f : A \rightarrow \mathbb{C} \mid \int_A |f(t)|^2 dt < \infty \right\}$$

is a Hilbert space for any measurable  $A$ , but our main interest is in  $L^2(\mathbb{R})$ .

**Theorem 5.**  $L^2(\mathbb{R})$  is a Hilbert space.

**Lemma 6.**  $L^2(\mathbb{R})$  is a vector space over the scalar field of  $\mathbb{C}$ .

*Proof.* Here we take the obvious addition and scalar multiplication  $(+) : L^2(\mathbb{R}) \times L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$  and  $(\cdot) : \mathbb{C} \times L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$  where  $(f + g)(x) = f(x) + g(x)$  and  $(\alpha \cdot f)(x) = \alpha f(x)$ .

We first show that these operations are well defined with respect to the equivalence classes of  $\sim$ . Let  $f_1 \sim f_2$  and  $g_1 \sim g_2$  be functions of  $L^2(\mathbb{R})$  equal except on measure zero sets  $Z_f$  and  $Z_g$  respectively. Clearly  $\alpha \cdot f_1(x) = \alpha \cdot f_2(x)$  except on  $Z_f$ , so  $\alpha \cdot f_1 \sim \alpha \cdot f_2$ . We have  $f_1(x) + g_1(x) = f_2(x) + g_2(x)$  except possibly on  $Z_f \cup Z_g$ , which is also of measure zero. Then  $f_1 + g_1 \sim f_2 + g_2$ .

We next demonstrate that these binary operations are closed on  $L^2(\mathbb{R})$ . The closure of  $L^2(\mathbb{R})$  under  $\cdot$  is implied by the linearity of the integral. We next observe that if  $f, g \in L^2(\mathbb{R})$  then

$$0 \leq \int_{\mathbb{R}} (|\bar{f}(x)| - |g(x)|)^2 dx = \int_{\mathbb{R}} |f(x)|^2 dx - 2 \int_{\mathbb{R}} |\bar{f}(x)g(x)| dx + \int_{\mathbb{R}} |g(x)|^2 dx$$

implies

$$0 \leq \int_{\mathbb{R}} |\bar{f}(x)g(x)| dx \leq \frac{1}{2} \int_{\mathbb{R}} |f(x)|^2 dx + \frac{1}{2} \int_{\mathbb{R}} |g(x)|^2 dx < \infty$$

then  $\bar{f}g \in L^2(\mathbb{R})$ . The closure of  $L^2(\mathbb{R})$  under  $+$  follows from

$$\begin{aligned} \int_{\mathbb{R}} |(f + g)(x)|^2 dx &= \int_{\mathbb{R}} \overline{(f + g)(x)} (f + g)(x) dx \\ &= \int_{\mathbb{R}} |f(x)|^2 dx + \int_{\mathbb{R}} \bar{f}g(x) dx + \int_{\mathbb{R}} f\bar{g}(x) dx + \int_{\mathbb{R}} |g(x)|^2 dx \\ &< \infty. \end{aligned}$$

The remaining vector space properties all follow from the definition and the fact that  $\mathbb{C}$  is a field.  $\square$

The map  $\langle \cdot | \cdot \rangle : L^2(\mathbb{R}) \times L^2(\mathbb{R}) \rightarrow \mathbb{C}$  where

$$\langle f | g \rangle = \int_{\mathbb{R}} \overline{f(x)} g(x) dx$$

defines an inner product on  $L^2(\mathbb{R})$ . We have from the linearity of the integral that  $\langle f | \alpha_1 g_1 + \alpha_2 g_2 \rangle = \alpha_1 \langle f | g_1 \rangle + \alpha_2 \langle f | g_2 \rangle$  for any  $\alpha_1, \alpha_2 \in \mathbb{C}$  and  $f, g_1, g_2 \in L^2(\mathbb{R})$ . Observe also that  $\langle g | f \rangle = \overline{\langle f | g \rangle}$ . We define the norm  $\|\cdot\| : L^2(\mathbb{R}) \rightarrow \mathbb{R}_0^+$  by

$$\|f\| = \langle f | f \rangle^{\frac{1}{2}} = \left( \int_{\mathbb{R}} |f(x)|^2 dx \right)^{\frac{1}{2}}.$$

Naturally  $\|f\| = 0$  if and only if  $f(x) = 0$  a.e.  $\mathbb{R}$ . The linearity of the integral implies that for any  $\alpha \in \mathbb{C}$  we have  $\|\alpha f\| = |\alpha| \|f\|$ . Observe further that

$$\begin{aligned} 0 &\leq \int_{\mathbb{R}} \int_{\mathbb{R}} |f(x)g(y) - f(y)g(x)|^2 dx dy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} |f(x)|^2 |g(y)|^2 + |f(y)|^2 |g(x)|^2 - \overline{f(x)}g(x)f(y)\overline{g(y)} - \overline{f(y)}g(y)f(x)\overline{g(x)} dx dy \\ &= 2\|f\|^2 \|g\|^2 - 2\langle f|g \rangle \langle g|f \rangle \\ &= 2\|f\|^2 \|g\|^2 - 2|\langle f|g \rangle|^2. \end{aligned}$$

With the final result showing the frequently useful Cauchy-Swartz inequality:

$$|\langle f|g \rangle| \leq \|f\| \|g\|.$$

The final property of the norm we are interested in is the triangle inequality. Observe that

$$\begin{aligned} \|f + g\| &= \left( \int_{\mathbb{R}} |f(x)|^2 + |g(x)|^2 + \overline{f(x)}g(x) + f(x)\overline{g(x)} dx \right)^{\frac{1}{2}} \\ &= (\|f\|^2 + \|g\|^2 + \langle f|g \rangle + \langle g|f \rangle)^{\frac{1}{2}} \\ &\leq (\|f\|^2 + \|g\|^2 + 2\|f\| \|g\|)^{\frac{1}{2}} \\ &= \|f\| + \|g\| \end{aligned}$$

Note that the triangle inequality, the obvious symmetry, and the positive values of  $d(f, g) = \|f - g\|$  make it a metric on  $L^2(\mathbb{R})$ , where  $d(f, g) = 0$  if and only if  $f \sim g$ .

We will require the following helpful result to show that this metric is Cauchy complete on  $L^2(\mathbb{R})$ .

**Lemma 7** (Monotone Convergence Theorem). *Let  $(f_n : \mathbb{R} \rightarrow \mathbb{R}_0^+ \cup \{\infty\} \mid n \in \mathbb{Z}^+)$  be a sequence of non-negative, measurable functions such that if  $i \leq j$  then  $f_i(x) \leq f_j(x)$  almost everywhere in  $\mathbb{R}$ . Then  $f : \mathbb{R} \rightarrow \mathbb{R}_0^+ \cup \{\infty\}$  defined almost everywhere by  $f(x) = \lim_{n \rightarrow \infty} f_n(x)$  is measurable and*

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}} f_n(x) dx = \int_{\mathbb{R}} f(x) dx.$$

*Proof.* We first claim that  $f$  is measurable. Let  $Z \subset \mathbb{R}$  be the set such that  $x \in Z$  if  $(f_n(x) \mid n \in \mathbb{Z}^+)$  is not monotone. By hypothesis,  $Z$  has 0 measure,  $f_n$  is measurable for every  $n \in \mathbb{Z}^+$ . This means that for every  $a \in \mathbb{R}$  we have  $\{x \in \mathbb{R} \mid f_n(x) > a\}$  is measurable. Because by Theorem 4 the  $\sigma$ -algebra of measurable sets is closed under complements, intersections, and unions, we see that  $\{x \in \mathbb{R} \mid f_n(x) > a\} \cap Z^c$  is measurable. In  $Z^c$  we see  $f(x) \geq f_n(x)$  for every  $n$ , so

$$\{x \in Z^c \mid f(x) > a\} \supseteq \bigcup_{n \in \mathbb{Z}^+} \{x \in Z^c \mid f_n(x) > a\}$$

We will see the reverse inclusion is also true. We have in  $Z^c$  that  $f(x) = \sup\{f_n(x) \mid n \in \mathbb{Z}^+\}$ . Then if  $f(x) > a$  there must be some  $n$  with  $f_n(x) > a$ , or  $a$  would be an upper bound of  $\{f_n(x) \mid n \in \mathbb{Z}^+\}$ . Hence

$$\{x \in Z^c \mid f(x) > a\} = \bigcup_{n \in \mathbb{Z}^+} \{x \in Z^c \mid f_n(x) > a\}$$

so  $\{x \in Z^c \mid f(x) > a\}$  is measurable for every  $a \in \mathbb{R}$ . As  $\{x \in Z \mid f(x) > a\} \subseteq Z$  we have that  $\{x \in Z \mid f(x) > a\}$  is measurable, hence

$$\{x \in \mathbb{R} \mid f(x) > a\} = \{x \in Z^c \mid f(x) > a\} \cup \{x \in Z \mid f(x) > a\}$$

is measurable for every  $a \in \mathbb{R}$ . Then  $f$  is a measurable function.

Observe that  $(\int_{\mathbb{R}} f_n(x) dx \mid n \in \mathbb{Z}^+)$  is a monotone sequence (possibly infinity) bounded above by  $\int_{\mathbb{R}} f(x) dx$  (possibly infinity), we have that  $(\int_{\mathbb{R}} f_n(x) dx \mid n \in \mathbb{Z}^+)$  must converge to some value  $\alpha \in \mathbb{R} \cup \{\infty\}$

$$\alpha = \lim_{n \rightarrow \infty} \int_{\mathbb{R}} f_n(x) dx \leq \int_{\mathbb{R}} f(x) dx.$$

If  $\alpha = \infty$  or  $\int_{\mathbb{R}} f_n(x) dx = \infty$  for any  $n \in \mathbb{Z}^+$ , then  $\lim_{n \rightarrow \infty} \int_{\mathbb{R}} f_n(x) dx = \int_{\mathbb{R}} f(x) dx$ . Otherwise, let  $s$  be any simple, measurable function with  $0 \leq s(x) \leq f(x)$  in  $Z^c$ , and fix some  $\beta \in (0, 1)$ . Let  $A_n = \{x \in Z^c \mid \beta s(x) \leq f_n(x)\}$ . Because  $(f_n(x) \mid n \in \mathbb{Z}^+)$  is monotone in  $Z^c$ , it follows that  $A_n \subseteq A_{n+1}$ . For any  $x \in Z^c$ , either  $f(x) = f_1(x) = \beta s(x) = 0$  and  $x \in A_1$  or  $f(x) > \beta s(x)$  and there is some  $n \in \mathbb{Z}^+$  such that  $f_n(x) \geq \beta s(x)$ . So  $Z^c = \bigcup_{n \in \mathbb{Z}^+} A_n$ . We see that  $\beta s(x) \leq f(x)$  implies

$$\int_{A_n} \beta s(x) dx \leq \int_{A_n} f_n(x) dx \leq \int_{\mathbb{R}} f_n(x) dx \quad (4.5)$$

We take the limit as  $n \rightarrow \infty$  on both sides of equation (4.5). Because the integral is a countably additive set function and  $(A_n \mid n \in \mathbb{Z}^+)$  non-decreasing we have that

$$\lim_{n \rightarrow \infty} \int_{A_n} \beta s(x) dx = \int_{\bigcup_{n \in \mathbb{Z}^+} A_n} \beta s(x) dx$$

And as  $\mathbb{R}$  and  $\bigcup_{n \in \mathbb{Z}^+} A_n$  differ only by  $Z$  we have

$$\int_{\mathbb{R}} \beta s(x) dx = \int_{\bigcup_{n \in \mathbb{Z}^+} A_n} \beta s(x) dx$$

Recall that the integral is linear, and equation (4.5) becomes

$$\beta \int_{\mathbb{R}} s(x) dx \leq \lim_{n \rightarrow \infty} \int_{\mathbb{R}} f_n(x) dx.$$

As  $\beta$  is arbitrary we may take  $\beta \rightarrow 1$  to obtain

$$\int_{\mathbb{R}} s(x) dx \leq \lim_{n \rightarrow \infty} \int_{\mathbb{R}} f_n(x) dx.$$



By taking the supremum over all non-negative simple functions less than  $f$  almost everywhere, we arrive at the definition of the integral on the left so

$$\int_{\mathbb{R}} f(x) dx \leq \lim_{n \rightarrow \infty} \int_{\mathbb{R}} f_n(x) dx.$$

□

**Theorem 8.**  $L^2(\mathbb{R})$  is Cauchy complete.

*Proof.* Let  $(f_n)$  be a Cauchy sequence of functions in  $L^2(\mathbb{R})$ . As the distance between consecutive functions goes to zero we can find a subsequence  $(f_{n_k})$  such that  $\|f_{n_k} - f_{n_{k+1}}\| < \frac{1}{2^k}$ . Then for any  $g \in L^2(\mathbb{R})$  we have by the Cauchy-Swarz inequality that

$$|\langle g | f_{n_k} - f_{n_{k+1}} \rangle| < \frac{\|g\|}{2^k}$$

so we see that

$$\begin{aligned} \|g\| &> \sum_{k=1}^{\infty} |\langle g | f_{n_k} - f_{n_{k+1}} \rangle| \\ &= \sum_{k=1}^{\infty} \int_{\mathbb{R}} |g(x)| |f_{n_k}(x) - f_{n_{k+1}}(x)| dx \\ &= \int_{\mathbb{R}} |g(x)| \sum_{k=1}^{\infty} |f_{n_k}(x) - f_{n_{k+1}}(x)| dx. \end{aligned}$$

by the Monotone Convergence Theorem. But this holds for arbitrary  $g \in L^2(\mathbb{R})$ . Let  $[a, b] \subseteq \mathbb{R}$  and let  $g$  be the indicator on the closed interval  $[a, b]$ . As  $\|g\| = \sqrt{b-a} > \int_{[a,b]} \sum_{k=1}^{\infty} |f_{n_k}(x) - f_{n_{k+1}}(x)| dx$  it must be that  $\sum_{k=1}^{\infty} |f_{n_k}(x) - f_{n_{k+1}}(x)|$  is infinite almost everywhere on  $[a, b]$ . As  $[a, b]$  is arbitrary, it must be that  $\sum_{k=1}^{\infty} |f_{n_k}(x) - f_{n_{k+1}}(x)|$  is finite almost everywhere on  $\mathbb{R}$ . If  $\sum_{k=1}^{\infty} |f_{n_k}(x) - f_{n_{k+1}}(x)|$  converges, then for any  $\varepsilon > 0$  it must be there is some  $N \in \mathbb{Z}^+$  such that

$$\sum_{k=N}^{\infty} |f_{n_k}(x) - f_{n_{k+1}}(x)| < \varepsilon,$$

so if  $i, j \geq N$  we have

$$|f_{n_i}(x) - f_{n_j}(x)| \leq \sum_{k=i}^{j-1} |f_{n_k}(x) - f_{n_{k+1}}(x)| < \varepsilon,$$

So for almost every  $x \in \mathbb{R}$ ,  $(f_{n_k}(x))_{k \in \mathbb{Z}^+}$  is a Cauchy sequence in  $\mathbb{R}$ . Then by the Cauchy completeness of  $\mathbb{R}$  we can define  $f(x) = \lim_{n_k \rightarrow \infty} f_{n_k}(x)$  whenever it converges (pointwise) and  $f(x) = 0$  everywhere else (which is at most some measure zero set). We claim that  $f_n \rightarrow f$  in  $L^2(\mathbb{R})$ .

Fix some  $\varepsilon > 0$ . We have that there is some  $N \in \mathbb{Z}^+$  with the property that if  $n, m \geq N$  then  $\|f_m - f_n\| < \varepsilon$ . Let  $n_k > N$ . Consider  $g_j(x) = \inf\{|f_{n_i}(x) - f_{n_k}(x)|^2 \mid i \geq j\}$ .

$$\begin{aligned}\|f - f_{n_k}\| &= \left( \int_{\mathbb{R}} |f(x) - f_{n_k}(x)|^2 dx \right)^{\frac{1}{2}} \\ &= \left( \int_{\mathbb{R}} \lim_{j \rightarrow \infty} g_j(x) dx \right)^{\frac{1}{2}}.\end{aligned}$$

We have for almost every  $x$  in  $\mathbb{R}$  that  $0 \leq g_1(x) \leq g_2(x) \leq \dots \leq |f(x) - f_{n_k}(x)|^2$  and  $g_j(x) \rightarrow |f(x) - f_{n_k}(x)|^2$  so by the Monotone Convergence Theorem we have:

$$\begin{aligned}\|f - f_{n_k}\| &= \left( \lim_{j \rightarrow \infty} \int_{\mathbb{R}} g_j(x) dx \right)^{\frac{1}{2}} \\ &= \left( \lim_{i \rightarrow \infty} \int_{\mathbb{R}} \inf\{|f_{n_i}(x) - f_{n_k}(x)|^2 \mid i \geq j\} dx \right)^{\frac{1}{2}} \\ &\leq \left( \lim_{i \rightarrow \infty} \inf_{i \geq k} \int_{\mathbb{R}} |f_{n_i}(x) - f_{n_k}(x)|^2 dx \right)^{\frac{1}{2}} \\ &\leq \varepsilon.\end{aligned}$$

Then  $f - f_{n_k} \in L^2(\mathbb{R})$  and as  $f_{n_k} \in L^2(\mathbb{R})$  we have by the additive closure of  $L^2(\mathbb{R})$  that  $f \in L^2(\mathbb{R})$ . We also see that for any  $n > N$  we have there is some  $n_k > n$  as defined above such that  $\|f - f_n\| \leq \|f - f_{n_k}\| + \|f_{n_k} - f_n\| < 2\varepsilon$ . As  $\varepsilon$  is arbitrary it must be that  $f_n \rightarrow f$ .  $\square$

# Bibliography

- [1] Charles K. Chui. *An introduction to wavelets*. Academic Press Professional, Inc., San Diego, CA, USA, 1992.
- [2] Michael Frazier. *An introduction to wavelets through linear algebra*. Springer, 1999.
- [3] Walter Rudin. *Principles of mathematical analysis*. McGraw-Hill Book Co., New York, third edition, 1976. International Series in Pure and Applied Mathematics.
- [4] Terence Tao. *An epsilon of room. I: Real analysis. Pages from year three of a mathematical blog*. Graduate Studies in Mathematics 117. Providence, RI: American Mathematical Society (AMS) , 2010.