THE WONDERFUL WORLD OF WAVELETS

WILLIAM K. STAFFORD & VAHE K. VARTAN

1. History of Wavelets

Wavelets are essential in signal and image processing that we experience every day. When we view images on our computer screen, listen to electronic sound waves or search for finger prints over the FBI network, it is wavelets that make it all possible. The FBI digitizes each finger print using pixels and color assignments. With this great deal of data, the FBI employs the services of wavelets to compress the data to make it easy to save and transfer.[1]

While wavelets have a century long history in mathematics, it was only in the 1980's that the term wavelet was applied to the mathematics that today is considered the foundation of the study. Up until the 1980's the study of this math was called "atomic decompositions". The first basis of the idea of wavelets was in an appendix of a thesis written by A. Haar in 1909. From the 1920's to the 1970's, there continued to be mathematical theory developed behind wavelets, but there was little crossover with physicists or those looking for application. The challenges of image and signal processing that arose in he 1980's resulted in a new focus and innovation within the field of wavelets. There was little crossover or realization of the math behind the theory, however, which meant a great deal of work was duplicated. Wavelets were developed for specific applications and as a result, there are a great deal of different wavelets that apply to a multitude of areas. We will be focusing on the Haar wavelet for while it is not particularly applicable, it provides a basis for understanding the theory of wavelets and their intracacies. [1]

2. Motivation and Applications

[1] Consider $f, g \in L^2[0,1)$ such that

$$f(x) = \begin{cases} 1 & \text{if } x \in [0,.5) \\ -1 & \text{if } x \in [.5,1) \\ 0 & \text{otherwise} \end{cases}, \quad g(x) = 2x - 1.$$

Figure 1 displays the graphs of f and g. Figure 2 displays the graphs of the Fourier Series for f and g.

[1]Information in this section is derived from [6].

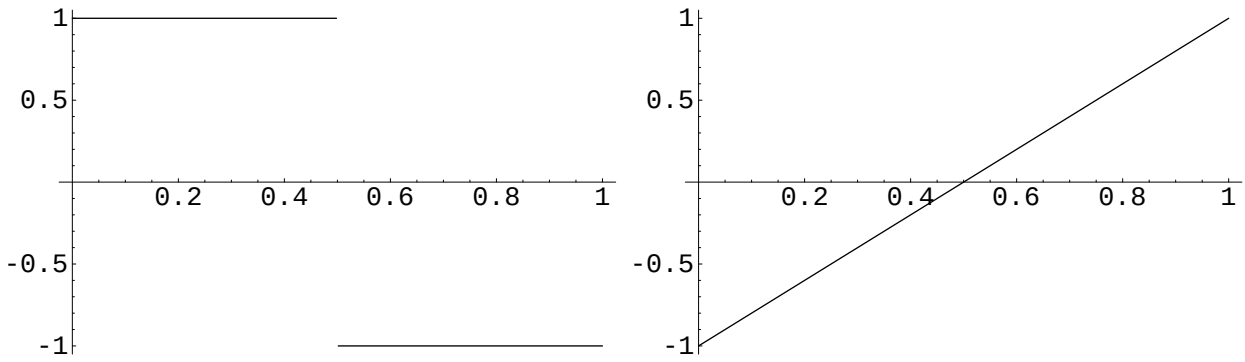

FIGURE 1. *Regular Functions f and g.*

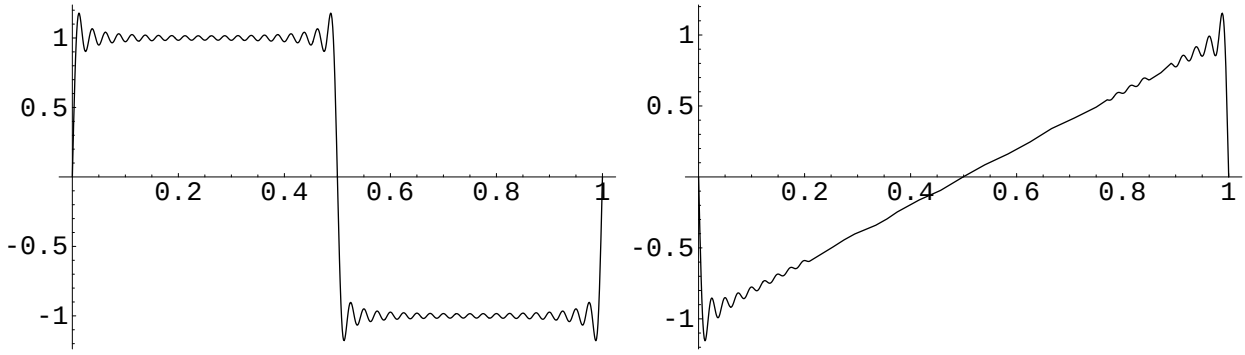

FIGURE 2. *Fourier Series Approximations.*

Figure 2 displays something known as "Gibb's phenomenon." This is witnessed at the discontinuities and edges of the graphs, at which point the Fourier series do not converge uniformly and so, it is evident that the Fourier series of f and g do not converge uniformly on $L^2[0,1)$. The function g is an example of a continuous function to which a Fourier series cannot converge uniformly. Hence, we would like a new basis that converges uniformly to *any* continuous function.

Now imagine an image on the computer screen. The computer screen is made up of tiny squares, known as pixels. To save an image, we might decide to save the color at each possible coordinate (pixel by pixel). For a relatively small resolution of 400x400 pixels, this would require us to save values at each of the $160,000$ coordinates. Saving each of these colors as a number would roughly require 8 bytes. Therefore, to save one image, it would require 1.28 megabytes, which is exorbitantly large for such a small resolution. Desirable resolutions produced by digital cameras are around 2592x1944 pixels; this would require 40.3 megabytes if we saved every coordinate.

If we think in a more clever fashion, we can see that an image can characterized by a function $f : \mathbb{R}^2 \to \mathbb{R}$, where our input is the (x, y) pixel coordinate, and the output

is the corresponding color value. It would make sense to approximate this function, because that way we could potentially be saving a lot of computer space. We could use Fourier series to approximate these functions, but if we think even harder, we would notice that images, for the most part, would be generated by a discontinuous function. Spots of black, next to spots of red, would cause many jumps across our color axis. The Fourier series is not even suited to converge uniformly to continuous functions in L^2, much less discontinuous functions. Gibb's phenomenon would severely hamper the way our images would be displayed. We would like to generate a new basis for L^2 to approximate these discontinuous functions such that they also approximate these functions "quickly". Hence, the production of the wavelet.

As stated above, there are many different wavelets for many different applications. MP3 audio files also employ the use of wavelets. The first use is to approximate the audio wave, which would decrease the amount of space needed to save an audio file. Yet another application is to smooth out noise, which is noticed as a series of sharp jumps on the graph of an audio wave. Noise causes unwanted and distorted sound; hence, the remastering of music is accomplished through the use of wavelets.

It is apparent from the above examples that wavelets take on a wide variety of digital uses.

3. The Haar Wavelet

[2] The goal of this section is to develop the Haar Wavelet. The Haar wavelet by itself is not useful but is imperative in the development of more advanced wavelet systems. In this section we will develop some of the key characteristics of the Haar wavelet, namely the Haar system is orthonormal on $[0, 1)$. We begin with the definition of a wavelet.

Definition 3.1. *A* **wavelet** *is a function $\psi \in L^2[0, 1)$ such that the set*

$$\{\psi_{kj}(x) = 2^{k/2}\psi(2^k x - j) : j, k \in \mathbb{Z}\}$$

forms an orthonormal basis for $L^2(\mathbb{R})$. ψ is equivalently referred to as the **mother wavelet**.

Definition 3.2. *The* **Haar system** *for $L^2[0, 1)$ is based off the mother wavelet where $\varphi = \chi_{[0,1)}$ and $\psi = \chi_{[0,.5)} - \chi_{[.5,1)}$. The Haar system is the family*

$$\{\varphi, \psi_{kj} : k \geq 0 \, and \, 0 \leq j < 2^k\}.$$

Example 1. *Figure 3.*

Lemma 3.3. The Haar system is orthonormal

[2]Information in this section is derived from [3]. Any derived proofs are cited at their ends.

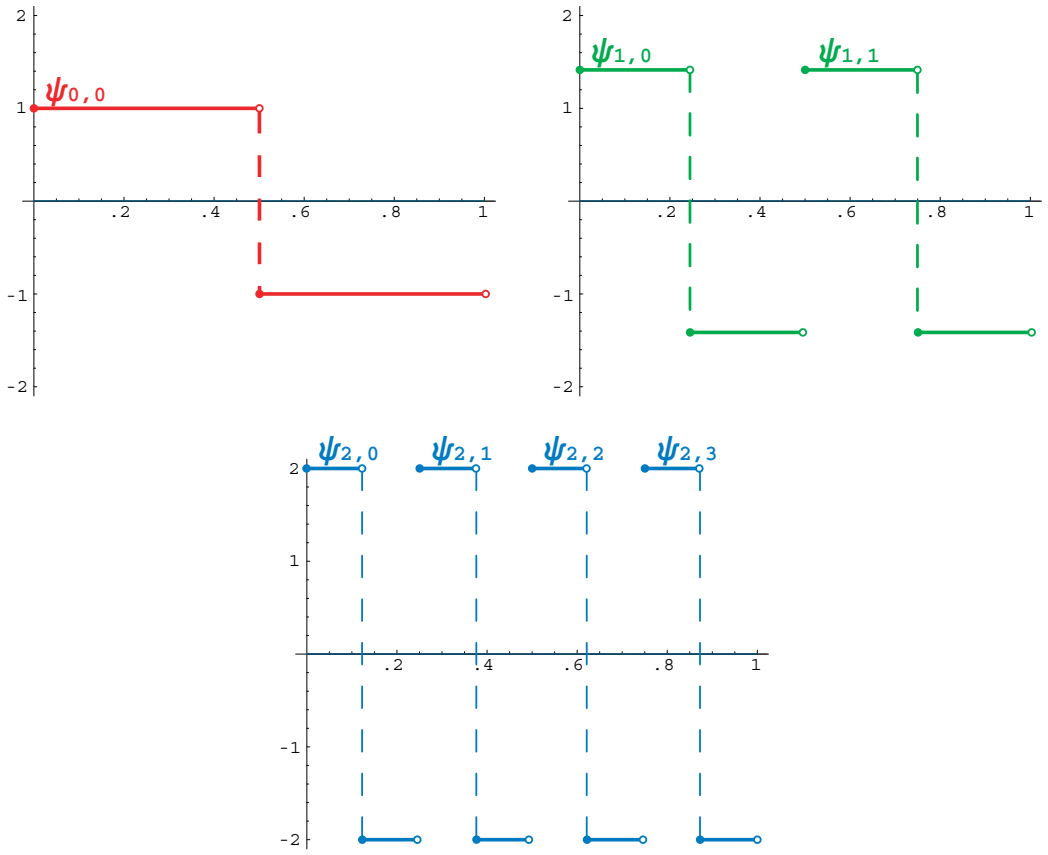

FIGURE 3. *Haar system for all $\psi_{k,j}$ where $k \leq 2$.*

Before we prove this, note that the inner product of $L^2[0,1)$ is defined to be $\int_{[0,1)} fg\,dm$. At many points throughout this paper, we will work with the Riemann integral when we are working with Riemann integrable functions. We are able to do this because the Lebesgue Integral generalizes the Riemann Integral.

Proof:

First, we will show that each function has norm 1, with respect to the L^2 norm. We will show that $\langle \varphi, \varphi \rangle_2 = 1$ (and hence $||\varphi||_2 = \sqrt{\langle \varphi, \varphi \rangle_2} = 1$).

Note that, for any interval A,

$$\chi_A(x)^2 = \begin{cases} 1^2 & \text{if } x \in A \\ 0^2 & \text{if } x \notin A \end{cases} = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} = \chi_A(x).$$

Now, we can show that:

$$\begin{aligned}\langle \varphi, \varphi \rangle_2 &= \int_0^1 \varphi^2 \, dx \\ &= \int_0^1 \chi_{[0,1)}(x)^2 \, dx \\ &= \int_0^1 \chi_{[0,1)}(x) \, dx = 1.\end{aligned}$$

We also want to show that $\langle \psi_{kj}, \psi_{kj} \rangle_2 = 1$ (and hence $||\psi_{kj}||_2 = \sqrt{\langle \psi_{kj}, \psi_{kj} \rangle_2} = 1$).

$$\begin{aligned}&\langle \psi_{kj}, \psi_{kj} \rangle_2 \\ &= \int_0^1 (\psi_{kj}(x))^2 \, dx \\ &= \int_0^1 \left(2^{k/2} \left(\chi_{[0,.5)}(2^k x - j) - \chi_{[.5,1)}(2^k x - j)\right)\right)^2 \, dx \\ &= 2^k \int_0^1 \chi_{[0,.5)}(2^k x - j)^2 - \chi_{[0,.5)}(2^k x - j)\chi_{[.5,1)}(2^k x - j) + \chi_{[.5,1)}(2^k x - j)^2 \, dx.\end{aligned}$$

$\chi_{[0,.5)}(2^k x - j)\chi_{[.5,1)}(2^k x - j)) = 0$ since the the intervals of the characteristic functions are disjoint. And so, following from our integral above:

$$\begin{aligned}&= 2^k \int_0^1 \chi_{[0,.5)}(2^k x - j) + \chi_{[.5,1)}(2^k x - j) \, dx \\ &= 2^k \int_0^1 \chi_{[0,1)}(2^k x - j) \, dx.\end{aligned}$$

We know that $\chi_{[0,1)}(2^k x - j) = 1$ when

$$\begin{aligned}&2^k x - j \in [0, 1) \\ &\implies 0 \leq 2^k x - j < 1 \\ &\implies j 2^{-k} \leq x < (j+1) 2^{-k}.\end{aligned}$$

Now we can rewrite:

$$\begin{aligned}&2^k \int_0^1 \chi_{[0,1)}(2^k x - j) \, dx \\ &= 2^k \int_{j2^{-k}}^{(j+1)2^{-k}} 1 \, dx \\ &= 2^k \left(2^{-k}\right) = 1.\end{aligned}$$

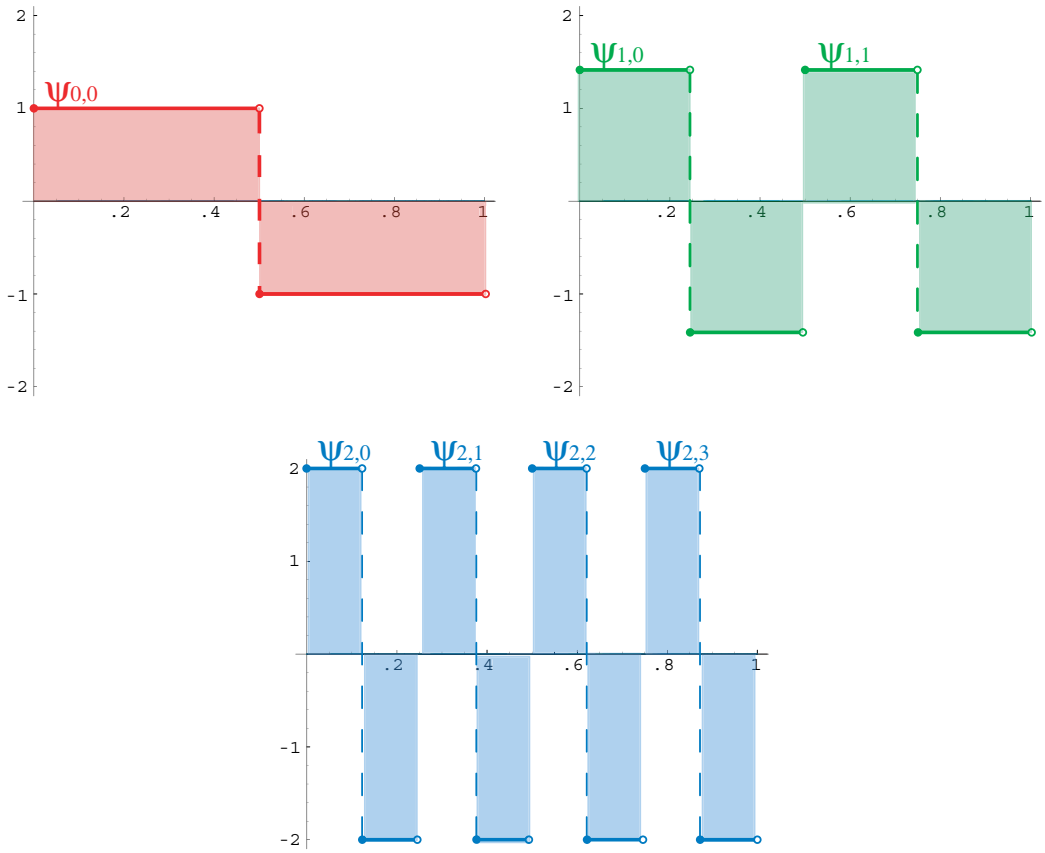

FIGURE 4. *Haar system for all $\psi_{k,j}$ where $k \leq 2$.*

Now we must show that the Haar system is orthogonal with respect to the L^2 norm, i.e. $\langle \varphi, \psi_{kj} \rangle_2 = 0$ and also, $\langle \psi_{kj}, \psi_{k'j'} \rangle_2 = 0$ where $k \neq k'$ and/or $j \neq j'$.

First we examine $\langle \varphi, \psi_{kj} \rangle_2$. Note that $\chi_A \chi_B = \chi_{A \cap B}$.

$$\begin{aligned}
\int_0^1 \varphi(x)\psi_{kj}(x)\,dx &= 2^k \int_0^1 \chi_{[0,1)}(x)\left(\chi_{[0,.5)}(2^k x - j) - \chi_{[.5,1)}(2^k x - j)\right)\,dx \\
&= 2^k \int_0^1 \chi_{[0,1)}(x)\chi_{[0,.5)}(2^k x - j) - \chi_{[0,1)}(x)\chi_{[.5,1)}(2^k x - j)\,dx \\
&= 2^k \int_0^1 \chi_{[0,.5)}(2^k x - j) - \chi_{[.5,1)}(2^k x - j)\,dx \\
&= 2^k \left(\int_0^1 \chi_{[0,.5)}(2^k x - j)\,dx - \int_0^1 \chi_{[.5,1)}(2^k x - j)\,dx \right) \\
&= 2^k \left(2^{-(k+1)} - 2^{-(k+1)}\right) = 0.
\end{aligned}$$

To prove that $\langle \psi_{kj}, \psi_{k'j'} \rangle_2 = 0$, we provide Figures 4 and 5. The actual proof is a tedious induction, full of confusing notation and to top it off, it provides no insight

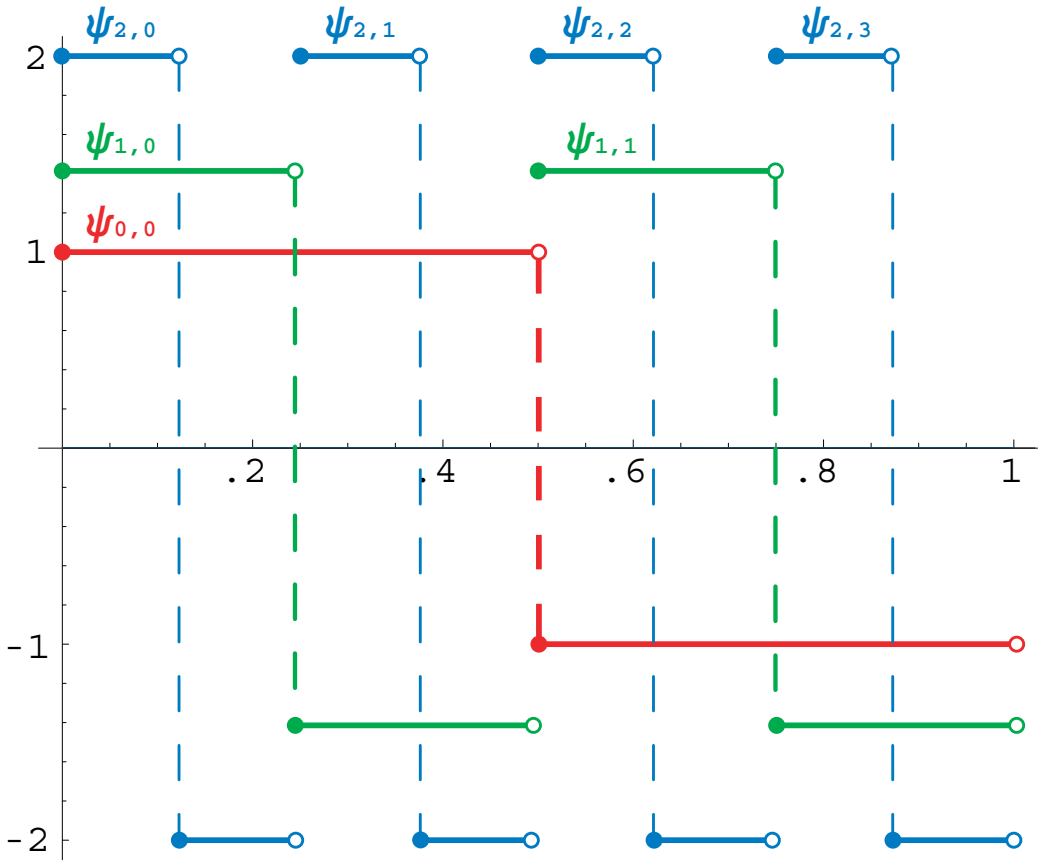

FIGURE 5. *Haar system for all* $\psi_{k,j}$ *where* $k \leq 2$.

into wavelet theory. A graphically motivated explanation is better for this reason. Looking at Figure 5, we notice that as k grows larger, the interval on which the Haar wavelet does not equal 0 is scaled by a factor of 2^{-1}. Heuristically, as k grows larger, the non-zero portion is squeezed along the x-axis and stretched along the y-axis. If we were to multiply ψ_{kj} by $\psi_{k'j'}$, where WLOG $k' > k$, the non-zero values of $\psi_{k'j'}$ would either be directly over a constant portion of ψ_{kj} or it would be completely disjoint. If $\psi_{k'j'}$ is over a constant portion, we would simply scale it by that value and so, $\psi_{k'j'}$ would still be symmetric about the x-axis; hence the inner product would be 0 (as displayed by Figure 4). If $\psi_{k'j'}$ is disjoint, then we would simply be multiplying the characteristic functions of disjoint sets, which we know also yields 0.

Therefore, we have shown that the Haar system is an orthonormal set.[3] □

Now that we have an orthonormal set, we can use the idea of inner product expansion to approximate a function, similar to what we did with our Fourier series approximations. We can represent this series approximation for the Haar System as follows:

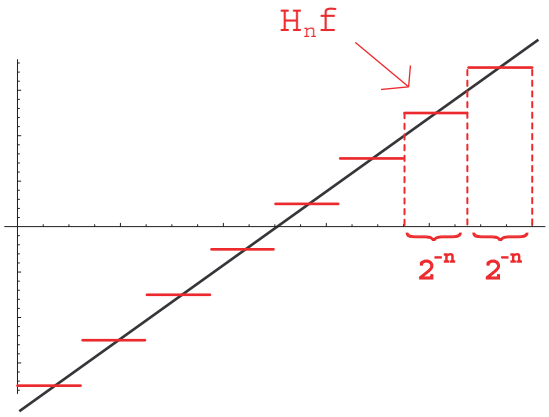

FIGURE 6. *Sketch of Lemma 3.6.*

Definition 3.4. *We denote the projection of* $f \in L^2[0,1)$ *with respect to the Haar System to be* $H_n f(x) := \langle f, \varphi\rangle\varphi(x) + \sum_{k=0}^{n-1}\sum_{j=0}^{2^k-1}\langle f, \psi_{kj}\rangle\psi_{kj}(x)$.

Proposition 3.5. Bessel's Inequality: If $\{e_i\}$ is an orthonormal set in a Hilbert Space H, then for any $x \in H$, $\sum\langle x, e_i\rangle^2 \leq ||x||^2$.

As a direct consequence of Bessel's Inequality, $\|H_n f\|_2 \leq \|f\|_2$. Now, all we need is one more tool to prove that the Haar System is an orthonormal basis.

Lemma 3.6. Let $f \in L^2[0,1)$ and denote $A_j^n = [j2^{-n}, (j+1)2^{-n})$, where $0 \leq j < 2^n, j \in \mathbb{Z}$. For $x \in A_j^n$,

$$H_n f(x) = 2^n \int_{A_j^n} f(t)\, dm.$$

A sketch of Lemma 3.6 is provided via Figure 6. The lemma is basically stating that $H_n f$ is constant and equal to the average value of the function f on each interval of length 2^{-n}.

Proof:
Note that $[0,1) = \bigcup_{j=0}^{2^n-1} A_j^n$ where each A_j^n is disjoint. We will prove by induction that

$$H_n f = \left(2^n \int_{A_j^n} f\, dm\right) \chi_{A_j^n}(x) \text{ for } x \in A_j^n.$$

For the base case $n = 1$, assume WLOG that $x \in A_0^1 = [0, \frac{1}{2})$ (the exact same method applies for $x \in A_1^1$). Then,

$$\begin{aligned} H_n f(x) &= \langle f, \varphi \rangle_2 \, \varphi + \langle f, \psi_{0,0} \rangle_2 \, \psi_{0,0} \\ &= \left(\int_{[0,1)} f \, dm \right) \chi_{[0,1)}(x) + \left(\int_{A_0^1} f \, dm - \int_{A_1^1} f \, dm \right) \left(\chi_{A_0^1}(x) - \chi_{A_1^1}(x) \right) \\ &= \left(\int_{A_0^1} f \, dm + \int_{A_1^1} f \, dm \right) \chi_{A_0^1}(x) + \left(\int_{A_0^1} f \, dm - \int_{A_1^1} f \, dm \right) \chi_{A_0^1}(x) \\ &= \left(2 \int_{A_0^1} f \, dm \right) \chi_{A_0^1}(x). \end{aligned}$$

Assume that

$$H_{n-1} f = \left(2^{n-1} \int_{A_j^{n-1}} f \, dm \right) \chi_{A_j^{n-1}}(x) \text{ for } x \in A_j^{n-1}.$$

Now we prove for $H_n f$. WLOG assume j is even. We can see that

$$\begin{aligned} H_n f &= H_{n-1} f + \sum_{j=0}^{2^n - 1} \langle f, \psi_{n-1,j} \rangle_2 \, \psi_{n-1,j} \\ &= H_{n-1} f + \ldots + 2^{n-1} \left(\int_{A_j^n} f \, dm - \int_{A_{j+1}^n} f \, dm \right) \left(\chi_{A_j^n}(x) - \chi_{A_{j+1}^n}(x) \right) + \ldots \end{aligned}$$

Given $x \in A_j^n$, we know that

$$A_j^n \cup A_{j+1}^n = [j 2^{-n}, (j+2) 2^{-n}) = [(2^{-1} j) 2^{-(n-1)}, ((2^{-1} j) + 1) 2^{-(n-1)}).$$

Since j is even, $2^{-1} j \in \mathbb{Z}$ and so, denoting $j' = 2^{-1} j$, we have that

$$A_j^n \cup A_{j+1}^n = A_{j'}^{n-1} \text{ and } A_j^n \subset A_{j'}^{n-1}.$$

Therefore, for $x \in A_j^n$,

$$\begin{aligned} H_{n-1} f(x) &= 2^{n-1} \left(\int_{j'}^{n-1} f \, dm \right) \chi_{A_{j'}^{n-1}}(x) \\ &= 2^{n-1} \left(\int_{A_j^n} f \, dm + \int_{A_{j+1}^n} f \, dm \right) \chi_{A_j^n}(x). \end{aligned}$$

Putting everything together, for $x \in A_j^n$

$$\begin{aligned}
H_n f(x) &= 2^{n-1}\left(\int_{A_j^n} f\,dm + \int_{A_{j+1}^n} f\,dm\right)\chi_{A_j^n}(x) \\
&\quad + 2^{n-1}\left(\int_{A_j^n} f\,dm - \int_{A_{j+1}^n} f\,dm\right)\left(\chi_{A_j^n}(x) - \chi_{A_{j+1}^n}(x)\right) \\
&= 2^{n-1}\left(\int_{A_j^n} f\,dm + \int_{A_{j+1}^n} f\,dm\right)\chi_{A_j^n}(x) \\
&\quad + 2^{n-1}\left(\int_{A_j^n} f\,dm - \int_{A_{j+1}^n} f\,dm\right)\chi_{A_j^n}(x) \\
&= \left(2^{n-1}\int_{A_j^n} f\,dm + 2^{n-1}\int_{A_j^n} f\,dm\right)\chi_{A_j^n}(x) \\
&= \left(2^{n}\int_{A_j^n} f\,dm\right)\chi_{A_j^n}(x).
\end{aligned}$$

Since $x \in A_j^n$, the characteristic function $\chi_{A_j^n}(x) = 1$ and so, we have

$$H_n f(x) = 2^n \int_{A_j^n} f\,dm.$$

□

Theorem 3.7. Let $f \in L^2[0,1)$. If $f \in C[0,1]$, then $H_n f$ converges uniformly to f. Furthermore, for any $f \in L^2[0,1)$, $H_n f$ converges to f with respect to the L^2 norm.

The key idea behind this proof will be that continuous functions are dense in L^2. Our strategy will be that if we can show that our Haar System projection can converge to any continuous function uniformly , then it can converge to any function in L^2 by proxy.

Proof:

We are given $f \in C[0,1]$. Since f is continuous on a compact interval, f is uniformly continuous on $[0,1]$. Therefore, for all $\epsilon > 0$, there exists a δ such that if $|x-t| < \delta$, then $|f(x) - f(t)| < \epsilon$ for all $x, t \in [0,1]$. We want to show that $H_n f(x)$ converges to $f(x)$ uniformly. Given $\epsilon > 0$, choose N such that $2^{-N} < \delta$ (i.e. $N > log_2(\frac{1}{\delta})$), where δ is given by uniform continuity. Note that we have chosen N independent of x. For

$n \geq N$ and $x \in [j2^{-n}, (j+1)2^{-n})$, apply Lemma 3.6 so that:

$$\begin{aligned}
|H_n f(x) - f(x)| &= \left| 2^n \int_{j2^{-n}}^{(j+1)2^{-n}} f(t)\, dt - f(x) \right| \\
&= \left| 2^n \int_{j2^{-n}}^{(j+1)2^{-n}} f(t)\, dt - 2^n \int_{j2^{-n}}^{(j+1)2^{-n}} f(x)\, dt \right| \\
&= \left| 2^n \int_{j2^{-n}}^{(j+1)2^{-n}} f(t) - f(x)\, dt \right| \\
&\leq 2^n \int_{j2^{-n}}^{(j+1)2^{-n}} |f(t) - f(x)|\, dt \\
&< 2^n \int_{j2^{-n}}^{(j+1)2^{-n}} \epsilon\, dt = \epsilon.
\end{aligned}$$

Since we chose N independently of x, the convergence of $H_n f$ to f is uniform. Furthermore, if a sequence of functions converges to a function uniformly, then it also converges to that same function with respect to the L^2 norm.

Now assume we are given any $f \in L^2[0,1)$. Since the continuous functions are dense in L^2 (with respect to the L^2 norm), we can find a sequence of functions $(g_k) \in C[0,1]$ such that for all $\epsilon > 0$, there exists K such that if $k \geq K$, then $\| g_k - f \|_2 < \epsilon/3$. Moreover, since $(g_k) \in C[0,1]$, we know by our previous work that there exists N such that if $n \geq N$, then $\| H_n g_k - g_k \|_2 < \epsilon/3$.
Also, via the use of bilinearity of the inner product,

$$\begin{aligned}
H_n f - H_n g_k &= \langle \varphi, f \rangle_2 \varphi - \langle \varphi, g_k \rangle_2 \varphi + \sum_{k=0}^{n-1} \sum_{j=0}^{2^k-1} \langle \psi_{kj}, f \rangle_2 \psi_{kj} - \langle \psi_{kj}, g_k \rangle_2 \psi_{kj} \\
&= \left(\langle \varphi, f \rangle_2 - \langle \varphi, g_k \rangle_2 \right) \varphi + \sum_{k=0}^{n-1} \sum_{j=0}^{2^k-1} \left(\langle \psi_{kj}, f \rangle_2 - \langle \psi_{kj}, g_k \rangle_2 \right) \psi_{kj} \\
&= \langle \varphi, f - g_k \rangle_2 \varphi + \sum_{k=0}^{n-1} \sum_{j=0}^{2^k-1} \langle \psi_{kj}, f - g_k \rangle_2 \psi_{kj} \\
&= H_n \left(f - g_k \right).
\end{aligned}$$

Following from Bessel's Inequality is that $\| H_n (f - g_k) \|_2 \leq \| f - g_k \|_2$.

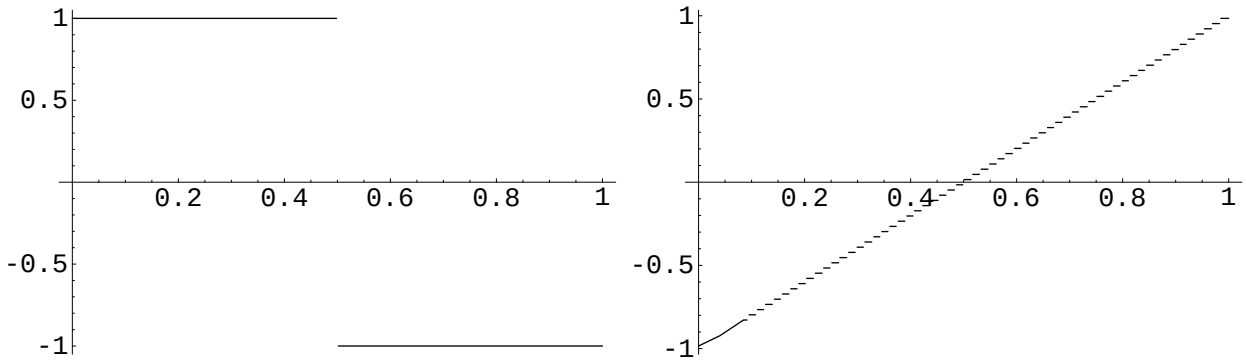

FIGURE 7. *Haar System approximations of f and g from Figure 1.*

Now, choose K and N accordingly from above so that:

$$\begin{aligned} \| H_n f - f \|_2 &\leq \| H_n f - H_n g_k \|_2 + \| H_n g_k - g_k \|_2 + \| g_k - f \|_2 \\ &\leq \| f - g_k \|_2 + \| H_n g_k - g_k \|_2 + \| g_k - f \|_2 \\ &< \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon. \end{aligned}$$

And so, we have shown that $H_n f$ converges to any $f \in L^2[0,1)$ with respect to the L^2 norm.[3] □

Corollary 3.8. The Haar system is an orthonormal basis for $L^2[0,1)$.

4. A Quick Application and the Shortcomings of the Haar System

We know the Haar System is a basis for $L^2[0,1)$, so it is possible to begin approximating functions with our projection $H_n f$. Figure 7 displays the Haar System approximation to the functions f and g from Figure 1. One can see that the Haar System displays the ability to converge to continuous functions extremely well. Figure 9 shows a Haar System approximation of $sin(2\pi x)$ generated by 64 terms. Figure 8 displays an imitation of what an audio wave might look like. The last approximation was generated by 512 terms, which is not very optimal since our audio wave is not very long or "noisy."

One should also notice that the Haar System is not good at remastering an audio wave; it will not smooth out sharp edges that cause audio "noise." There are other wavelets that are optimal for remastering that also converge more quickly.[3]

In conclusion, it is evident that the Haar System does not actually accomplish our needs in the real world. This is not to completely dismiss Haar wavelets as useless; Haar wavelets build the foundation for other wavelets. Most importantly, the shortcomings of the Haar System are able to be fixed. Our journey has only shown

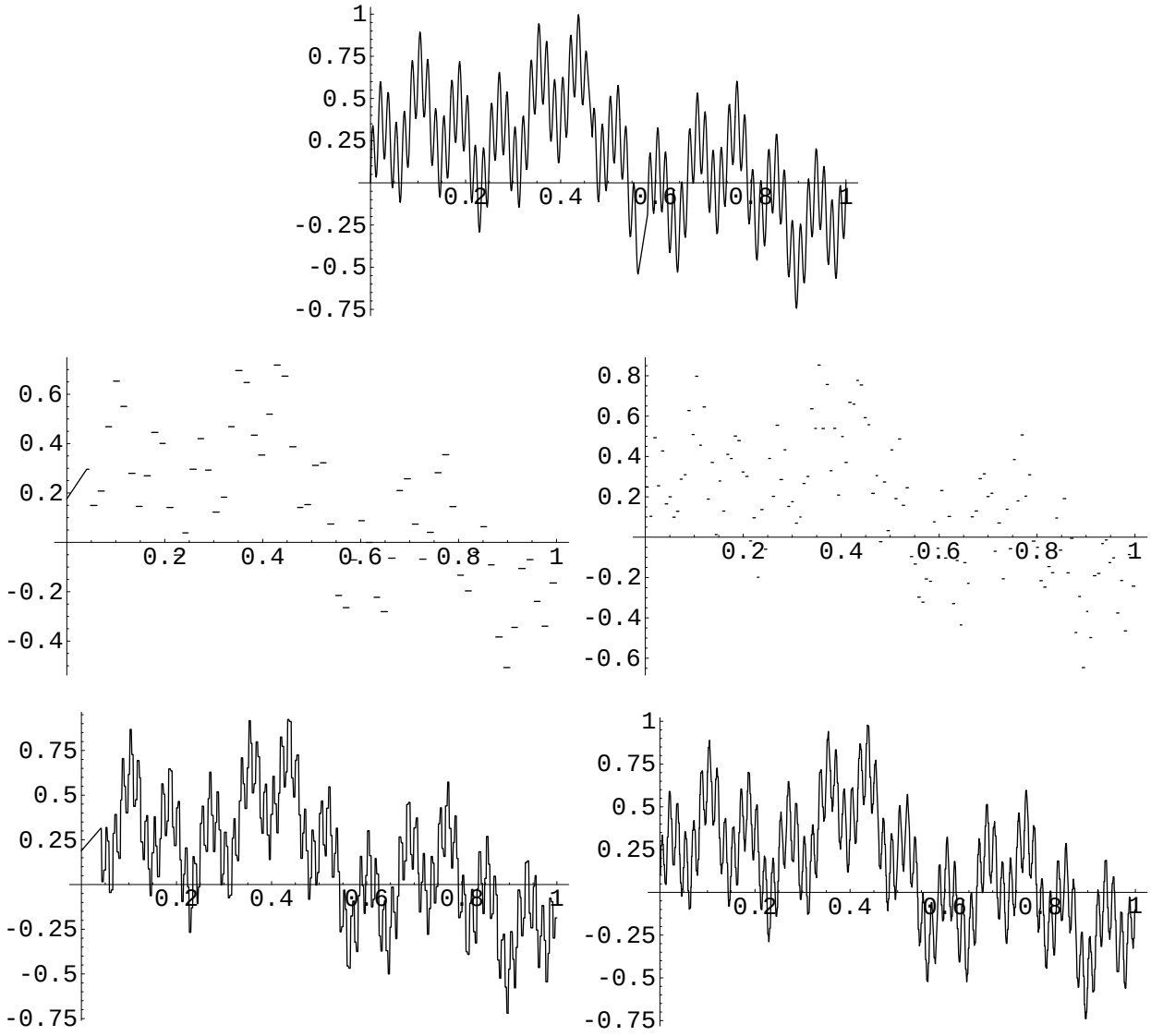

FIGURE 8. *An audio wave imitation followed by Haar System approximations (k bounded by 5, 6, 7, 8).*

us the entrance to wavelet theory, but we have equipped ourselves with the necessary tools to step through and immerse ourselves in the Wonderful World of Wavelets.

REFERENCES

[1] Aboufadel, Edward and Seven Schlicker. *Discovering Wavelets.* John Wiley & Sons, Inc. 1999 New York.

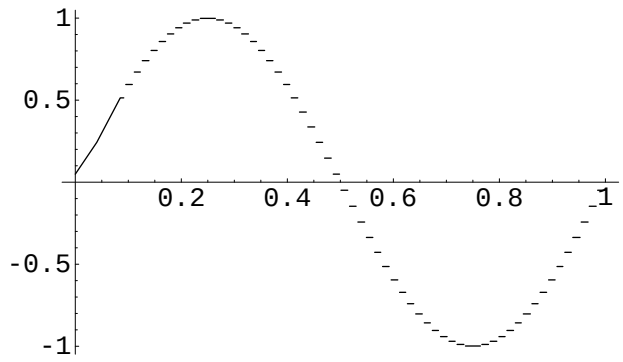

FIGURE 9. *Haar System approximation of* $sin(2\pi x)$ *(k bounded by 5).*

[2] Bratteli, Ola and Palle Jorgensen. *Wavelets Through a Looking Glass*. Birkhauser 2002 Boston

[3] Davidson, Kenneth R. and Allan P. Donsig. *Real Analysis with Real Applications*. Prentice Hall, Inc. 2002 Upper Saddle River

[4] Kaiser, Gerald. *A Friendly Guide to Wavelets*. Birkhauser, 1994 Boston.

[5] Meyer,Yves. *Wavelets, Algorithms & Applications*. Society for Industrial and Applied Mathematics, 1993 Philadelphia.

[6] Sabloff, Josh. Master of Disaster.