SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

A guided tour of Wavelet theory via the constructions of Multiresolution analyses

av

David Sermoneta

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

A guided tour of Wavelet theory via the constructions of Multiresolution analyses

David Sermoneta

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Salvador Rodriguez-Lopez

2025 - No K31

Abstract

We give an exposition of the theory leading up to, and including, the subject of wavelet analysis. Wavelets provide an elegant and practical way to decompose functions in a fashion similar to that of decomposing vectors in $\mathbb{R}^n$, the key difference being that the bases for most function spaces are not finite. The main goal for this thesis is to first develop the necessary theory for what essentially amounts to "doing linear algebra" in an infinite-dimensional context, and then apply that framework to the study of Wavelet analysis. To start off, we develop a more flexible notion of integration called the Lebesgue integral, and use it to construct normed spaces of functions, such as $L^2(\mathbb{R})$, the inner product space of square-integrable functions. To build towards Wavelet analysis, we define the Fourier transform, a powerful tool in functional analysis that lets us describe functions in terms of complex exponentials. Finally, we introduce the theory of multiresolution analyses, a central object in the study of Wavelets, as they let us construct wavelet bases of $L^2(\mathbb{R})$ by decomposing the space into nested subspaces corresponding to different levels of "detail". This provides a framework for analyzing how the components of a function are distributed across these levels. Using this theory, we construct the famous Haar and Shannon wavelets and their corresponding bases.

Sammanfattning

Vi presenterar i detta arbete teorin som leder upp till, samt utgör analysen om Wavelets. Det som gör wavelets intressanta är att de erbjuder ett elegant, samt praktiskt sätt att bryta ner funktioner på sätt som liknar hur man bryter ner vektorer i $\mathbb{R}^n$, med stora skillnaden att vi inte längre talar om det i kontextet av ändligt dimensionella rum. Vi uppnår detta mål genom att först utveckla teorin som tillåter oss att använda maskineriet vi känner till från linjär algebra i ett oändligtdimensionellt kontext. Till att börja med, utvecklar vi en mer flexibel teori om integration som heter Lebesgueintegralen, och sedan använder vi den för att konstruera ett flertal normerade funktionsrum såsom $L^2)\mathbb{R})$, inre produktrummet av funktioner vars kvadrat är integrerbart. För att motivera, samt komma närmare till Waveletteori, behandlar vi ett mycket användbart medel inom funktionella analysen, nämligen Fourier transformen, som låter en att beskriva funktioner i termer av komplexa exponentialfunktioner. Slutligen introducerar vi teorin om multiresolutionsanalyser, ett centralt objekt inom studiet av Wavelets, eftersom de låter oss konstruera waveletbaser för $L^2(\mathbb{R})$ genom att dekomponera rummet i nästlade delrum som motsvarar olika nivåer av "detalj". Detta ger ett ramverk för att analysera hur komponenterna i en funktion är fördelade över dessa nivåer. Med hjälp av denna teori konstruerar vi även de välkända Haar och Shannon-waveletsen och deras motsvarande baser.

Acknowledgements

I want to start off by thanking my family for their indispensable support throughout my life. I also want to thank all my friends outside of mathematics, especially Gab, Filippa, Frej, Jakob, Charlie, Indigo, Zurabi and Andrejs.

Thank you, Emma, for these three years we have spent together. I am immensely grateful to have met you, and for all the time we have spent together, both as studying companions and as friends.

Thank you to all the people that I've met through, or grown closer to because of, mathematics: All the classmates and professors that I've had the pleasure to interact with, all the authors of math books I've had the pleasure to read, and all the nice people on various internet forums that have been of tremendous help throughout the years. I would also like to give a special thanks to Jonas, Ottilia, and all the good people in the DGG Math Discord server.

And finally, I want to thank Salvador for being a great mentor, offering guidance and support throughout the writing of this thesis. I have thoroughly enjoyed our discussions in meetings, both on and off topic, and I don't think I could have made a better choice of supervisor.

Contents

1 Introduction

Finding elementary functions that make up other kinds of functions on an interval is something mathematicians have had an interest in since the days of Leibniz and Newton. A good example of this, is the case of smooth functions, where polynomials are amazing at approximating them, and given analycity, taking a limit on this approximation yields us a power series representing the entire function! However, this is approach suffers one clear restriction: almost no function encountered in the real world is smooth! A way around this restriction is treated in the study of Fourier analysis. Joseph Fourier had a few brilliant ideas, and one of them was to think of periodic functions as a series of trigonometric polynomials of the form $C_n e^{inx}$, where $n \in \mathbb{Z}$, and $C_n \in \mathbb{C}$. A more modern interpretationn of this, is to think of it as the set of trigonometric polynomials being dense in the vector space of periodic functions. However, a few questions might arise: Is this even actually possible? What other conditions do we need to put beyond periodicity on these functions? What about non-periodic functions? In the case we can do this, how do we decide which coefficient to choose? And finally, how are the coefficients determined? The theory has come a long way since Fourier, and the subtleties have thus increased, along with the theory's expressive power. Often, when working with bases, an inner product is desirable, as that allows us to project onto basis elements. Since our vectors in this case are functions, a natural inner product to consider is the integral.

In section 2 we develop the Lebesgue integral which will allow us to integrate functions under more relaxed conditions than the Riemann integral, as well as pass limits under the integral sign.

In section 3 we make use of the Lebesgue integral to define a whole class of function spaces called L^p-spaces. Out of these spaces, L^2 is of special interest because it allows us to project functions similar to as if they were vectors in $\mathbb{R}^n$. The theory of this is expanded upon in section 3.2.

In section 4 we present the Fourier transform for $L^1(\mathbb{R})$ as well as $L^2(\mathbb{R})$ as a formal tool to mainly help us prove certain theorems that will be useful in the upcoming chapters. The Fourier transform is an operator that describes functions in terms of complex polynomials. A helpful way to think about it is as a sort of generalized projection, giving us information about a function over a continuum of values, not just discrete ones.

Finally section 4 also serves as a foundation for section 5 about wavelets in the sense that wavelets are motivated by and solve some of the problems of the Fourier transform. Most importantly, we can (and will) construct orthonormal bases of $L^2(\mathbb{R})$ using wavelets. Circling back to the original idea of Fourier, but for non-periodic functions.

The theory leading up to this result of the thesis is based on the lecture notes of Bengt Ove Turesson [TUR15], which constructs wavelets out of an object called a multiresolution analysis. In the end we will have a method for constructing wavelet bases of $L^2(\mathbb{R})$ from a single function satisfying just a handful of properties. To help us in this section, several examples are presented with the help of the Haar wavelet.

2 Integrals, integrals, integrals

2.1 Problems with the Riemann integral

We will go over two main pathologies of the Riemann integral, the first being that if a function is unbounded, the Riemann integral is essentially useless. The second one is that we often want to interchange limits and integrals. Suppose we have a sequence of Riemann integrable functions $f_n : [a,b] \to \mathbb{R}$, where $\lim_{n\to\infty} f_n(x) = f(x)$ for some function f, at each point $x \in [a,b]$. Then we can't know for sure whether or not f is also Riemann integrable. As we will see in the following example.

Example 1. Let $f : [0,1] \to \mathbb{R}$ be defined as

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational,} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

We can show that f is not Riemann integrable by the following argument. For any subinterval I of $[0,1]$, since both the irrationals and rationals are dense in $\mathbb{R}$, we get that

$$\sup_I f = 1, \text{ and } \inf_I f = 0.$$

Therefore the upper and lower integrals of f can never coincide, and so f cannot be Riemann integrable. Now let (r_n) be a bijection from the natural numbers to $\mathbb{Q} \cap [0,1]$, and define

$$f_k = \begin{cases} 1 & \text{if } x \in \{r_1, \ldots, r_k\}, \\ 0 & \text{otherwise.} \end{cases}$$

Then $f_k(x) \to f(x)$ as $k \to \infty$, and it is easy to see that each f_k is Riemann integrable with

$$\int_0^1 f_k = \sum_{i=1}^{k} \int_{r'_{i-1}}^{r'_i} f_k = 0, \text{ where } r'_{i-1} < r'_i.$$

What we have witnessed here is the fact that the function defined as the pointwise limit of a sequence of Riemann-integrable functions is not necessarily Riemann-integrable.

The following section will resolve the problems posed by the Riemann integrals.

2.2 σ-Algebras and Measurability

Definition 2. Suppose X is a set and $\mathcal{S}$ is a set of subsets of X, that is, $S \subset \mathcal{P}(X)$. Then $\mathcal{S}$ is called a σ-algebra on X if

1. $\emptyset \in \mathcal{S}$;
2. if $E \in \mathcal{S}$, then $X \setminus E \in \mathcal{S}$;
3. if $E_1, E_2, \ldots$ is a sequence of elements of $\mathcal{S}$, then $\bigcup_{k=1} E_k \in \mathcal{S}$.

We call each element in $\mathcal{S}$ a measureable set, and the pair $(X, \mathcal{S})$ a measurable space.

Example 3 (Examples of σ-algebras)**.**

1. Clearly $\{\emptyset, X\}$ is a σ-algebra on X for any set X.
2. The powerset of a set X, $\mathcal{P}(X)$ is a σ-algebra on X.
3. The set $\mathcal{S} = \{\bigcup_{n \in K}(n, n+1] : K \subset \mathbb{Z}\}$ is a σ-algebra on $\mathbb{R}$. If $K = \emptyset$, then $\emptyset \in \mathcal{S}$. Suppose now that $E_1, E_2, \ldots\ldots$ is a sequence of elements of $\mathcal{S}$, where to each E_k, we can associate a $K_k \subset \mathbb{Z}$, such that
$$E_k = \bigcup_{n \in K_k} (n, n+1].$$
Then $K = \bigcup_k K_k \subset \mathbb{Z}$, and it is not so hard to see that $\bigcup_k E_k = \bigcup_{n \in K}(n, n+1] \in \mathcal{S}$. Now suppose $E \in \mathcal{S}$, and $E = \bigcup_{n \in K}(n, n+1]$.
$$\mathbb{R} \setminus E = \bigcap_{n \in K} \mathbb{R} \setminus (n, n+1] = \mathbb{R} \setminus \bigcup_{n \in K} (n, n+1] = \bigcup_{n \in \mathbb{Z}} (n, n+1] \setminus \bigcup_{n \in K} (n, n+1].$$
This is still of the form $\bigcup_{k \in K'}(k, k+1]$, where $K' = \mathbb{Z} \setminus K$, and so $\mathbb{R} \setminus E \in \mathcal{S}$, and we are done.

Definition 4. Let $B \subset P(X)$. Then we call the smallest σ-algebra containing B the σ-algebra generated by B. More formally, The smallest σ-algebra containing B can be written as
$$\bigcap_{\mathcal{S} \supset B} \mathcal{S}.$$

Example 5. Suppose X is a set and $A = \{\{x\} : x \in X\}$. Then the σ-algebra generated by A is the set of all subsets $E \subset X$, such that either E is countable, or $X \setminus E$ is countable.

Proof. Let $\mathcal{S} = \{E \subset X : E \text{ or } X \setminus E \text{ is countable.}\}$. It is well-known that a countable union of countable sets is countable, so for any sequence contained in $\mathcal{S}$ of countable elements, their union is also in $\mathcal{S}$. Therefore, suppose the sequence $E_1, E_2, \ldots$ contained in $\mathcal{S}$ contains some E_i that is uncountable, then clearly the union $\cup_k E_k$ is uncountable, but
$$X \setminus \left(\bigcup_k E_k \right) = \bigcap_k (X \setminus E_k)$$
will be countable, since $X \setminus E_i$ will be. Now clearly, A is contained in $\mathcal{S}$. From this it follows that $\mathcal{S}$ is a σ-algebra that contains A. Now to show $\mathcal{S}$ is the smallest such σ-algebra, we wish for every

$E \in \mathcal{S}$ to be in every σ-algebra that contains A. Let $\mathcal{F}$ be any σ-algebra containing A. Then $\mathcal{F}$ by definition contains every countable subset of X, and so we may assume $E \subset X$ is uncountable and in $\mathcal{S}$, but then $X \setminus E$ is countable, and therefore is in $\mathcal{F}$. So E must also be in $\mathcal{F}$ and we are done. ☺

Definition 6. Let $X = \mathbb{R}$, and let B the collection of open intervals of $\mathbb{R}$, then the elements of the σ-algebra generated by B are called *Borel sets*, and we use $\mathcal{B}$ to denote this σ-algebra.

2.3 Measurable functions

Definition 7. Suppose $\mathcal{S}$ is a σ-algebra on X. We call a function $f : X \to [-\infty, \infty]$ *measurable*, if for any $a \in \mathbb{R}$,

$$f^{-1}((a, \infty]) \in \mathcal{S}.$$

In other words, if for every a, the inverse image of f under (a, ∞) is a measurable set, then f is called measurable.

Example 8. Let $E \subset X$. We denote the characteristic function of E by

$$\chi_E(x) = \begin{cases} 1 & \text{if } x \in E, \\ 0 & \text{if } x \notin E. \end{cases}$$

In the case that $X = \mathbb{R}$, χ_E is a measurable function if and only if E is measurable. Note that for any $a \in \mathbb{R}$,

$$\chi_E^{-1}((a, \infty]) = \begin{cases} E & \text{if } 0 \le a < 1, \\ X & \text{if } a < 0, \\ \emptyset & \text{if } 1 \le a. \end{cases}$$

So if E is not measurable, $\chi_E^{-1}((\frac{1}{2}, \infty])$ cannot be measurable, and if E is measurable, then the above shows that $\chi_E^{-1}((a, \infty])$ is always measurable.

Lemma 1. Let $\mathcal{S}$ be a σ-algebra on a set X, and $\{f_n\}_{n \in \mathbb{N}}$ a sequence of measurable functions from X to $[-\infty, \infty]$, then the functions defined pointwise at each $x \in X$ as

$$g(x) = \inf\{f_n(x) : n \in \mathbb{N}\} \text{ and } h(x) = \sup\{f_n(x) : n \in N\}$$

are both measurable functions.

Proof. Let $a \in \mathbb{R}$. If $h(x) \in (a, \infty]$, for some $x \in \mathbb{R}$, then by how $h(x)$ is defined, for some $n \in \mathbb{N}$, $a < f_n(x) \le h(x)$, and so $f_n(x) \in (a, \infty]$. For any $n \in \mathbb{N}$, there exists an x such that $f_n(x) \in (a, \infty]$, so $h(x) \ge f_n(x) > a$ implies $h(x) \in (a, \infty]$. This shows that $h^{-1}((a, \infty]) = \bigcup_{n \in \mathbb{N}} f_n^{-1}((a, \infty])$. Since each f_n is measurable, and a countable union of measurable sets is measurable, h must be measurable. For g, remember the relation

$$\inf\{f_n(x) : n \in \mathbb{N}\} = -\sup\{-f_n(x) : n \in \mathbb{N}\}$$

from real analysis, and so the result for g automatically follows from that. ☺

2.4 The Lebesgue measure

Now we will define the concept of a *measure* on a σ-algebra. We note here that the mathematical theory of measures is a lot more vast than how it is presented here.

Definition 9. Let $\mathcal{S}$ be a σ-algebra on a set X. A function $\mu : \mathcal{S} \to [0, \infty]$ is called a *measure* on $\mathcal{S}$ if $\mu(\emptyset) = 0$, and for any sequence of disjoint measurable sets sets $\{S_n\}_{n \geq 1}$,

$$\mu\left(\bigcup_{n=1}^{\infty} S_n\right) = \sum_{n=1}^{\infty} \mu(S_n).$$

This last property is called *countable additivity*.

Remark 1. Notice that it follows from μ being countably additive that if $A \subset B$, then $\mu(A) \leq \mu(B)$.

We would like to have a measure on the collection of Borel sets. However, it turns out that showing a function is a measure on a σ-algebra is hard, so instead we assume that the function $\ell : B \to [0, \infty]$ defined by

$$\ell((a,b)) = \begin{cases} b - a & \text{if } b, a \in \mathbb{R} \\ +\infty & \text{if } a = -\infty \text{ or } b = +\infty \end{cases}$$

can be extended to an measure $\mu : \mathcal{B} \to [0, \infty]$ on $\mathcal{B}$ via

$$\mu(E) = \inf\left\{\sum_{n=1}^{\infty} \ell(I_n) : E \subset \bigcup_{n=1}^{\infty} I_n\right\},$$

where each I_n is an open interval of $\mathbb{R}$. We will not show that μ actually satisfies countable additivity. Furthermore, we note that for any set $A \subset \mathbb{R}$ and $t \in \mathbb{R}$,

$$\mu(\{a + t : a \in A\}) = A.$$

This property just says that μ is invariant under translations. One might ask themselves why a measure couldn't just be defined in a similar fashion, but on all the subsets of $\mathbb{R}$ instead. After all, $\mathcal{P}(\mathbb{R})$ is itself a σ-algebra. It turns out that things cannot be so simple (we are dealing with the real numbers after all). One can actually show that there exists no non-negative real valued function, defined on the subsets of $\mathbb{R}$, that is both a measure on this $\mathcal{P}(\mathbb{R})$, and satisfies translation invariance. The reader is referred to [AXL20, p. 25] for a proof of this.

Definition 10. The measure μ defined above is called the *Lebesgue measure*.

2.5 Sets of measure 0 and on the notion of "Almost Everywhere"

One important class of Borel sets are the sets $A \subset \mathcal{B}$ such that $\mu(A) = 0$. There are many such sets, and most commonly, they are the countable (including finite) subsets of $\mathbb{R}$. Note that these are always Borel sets, since any singleton $\{x\}$ is a closed set, and hence its complement is a union of open intervals, another Borel set. Since $\{x\}$ is the complement of a Borel set, it's a Borel set. Since any countable set is a countable union of singletons, any countable set is a Borel set. To show that they all have measure 0, we again start with the singletons. Define for $x \in \mathbb{R}$, the sequence $\{I_n\}_{n\geq 0} = (x - \frac{1}{n}, x + \frac{1}{n})$. Then for all $n \geq 0$, $\mu(I_n) = \frac{2}{n}$, and

$$\mu(\{x\}) \leq \frac{2}{n}, \text{ by countable additivity,}$$

meaning that $\mu(\{x\}) = 0$. By extension, if A is a countable set $\mu(A) = 0$.

Remark 2. There are also uncountable Borel sets which have Lebesgue measure 0, like the Cantor set. But we shall leave this here for now. If the reader would like to know more, see [AXL20, p. 57].

Let $f : X \to [-\infty, \infty]$ be a measurable function. We say a property P of f holds almost everywhere (abbreviated to a.e.), if the subset $A \subset X$ for which the P of f does not hold has $\mu(A) = 0$. If g is also a measurable function, then define the relation $\sim$, via $f \sim g$ if and only if for $N = \{x \in X : f(x) \neq g(x)\}$ it holds that $\mu(N) = 0$. We can show that $\sim$ is an equivalence relation. Symmetry and reflexivity are fairly obvious. For transitivity, suppose $f \sim g$, and $g \sim h$, and define $Z = \{x \in X : f \neq h\}$. Then for any $x \in \mathbb{Z}$, $f(x) \neq g(x)$, or $g(x) \neq h(x)$, since otherwise they would both have to be equal, meaning that $f(x) = h(x)$, a contradiction. From this it follows that Z is contained in the union of two sets that have measure 0, and so Z must have measure 0 (by the additivity of measures).

Something peculiar is that you can have a function f that is almost everywhere equal to a continuous function g, but such that f itself is not continuous everywhere. An example of this is the characteristic function on the rationals, $\chi_{\mathbb{Q}}(x)$.

2.6 Integration of measurable functions

Now we are ready to define the a notion of integrability for measurable functions. First we want to define the integral for simple functions, and after that, make use of that simpler definition to define the integral for more complicated functions.

Definition 11. Let $(X, \mathcal{S})$ be a measurable space. A function $e : X \to \mathbb{R}$ is called a simple function, if it can be written as

$$e(x) = \sum_{i=1}^{n} \alpha_i \chi_{S_i}(x),$$

where $S_1, S_2, \ldots, S_n$ are measurable disjoint sets, and $\alpha_1, \ldots, \alpha_n$ are real numbers.

An immediate question arises: how can we make sure that simple functions are even enough to define integrals of any measurable function? The following theorem tells us that the simple functions are good point-wise approximations of measurable functions, and further, that this approximation is monotone.

Theorem 2. Let $(X, \mathcal{S})$ be a measurable space. Then for any non-negative measurable function $f : X \to [0, \infty]$, there exists an increasing sequence of positive simple functions $e_n : X \to [-\infty, \infty]$, such that for each $x \in X$,

$$\lim_{n\to\infty} e_n(x) = f(x).$$

Proof. Define e_n as follows:

$$e_n(x) = \begin{cases} 2^{-n}k & \text{if } 2^{-n}k \le f(x) < 2^{-n}(k+1), \\ & \quad \text{where } k = 0, 1, \ldots, 2^{2n} - 1, \\ 2^n & \text{if } 2^n \le f(x). \end{cases}$$

It's easy to verify that e_n is indeed a simple function, the important part being that e_n can at most take on 2^{2n} different (positive) values. Further, the sequence is increasing, since if $2^{n+1} \le f(x)$, then clearly $e_n \le e_{n+1}$. On the other hand, if for some n, $2^n \le f(x) \le 2^{n+1}$, then choosing $k = 2^{2n+1}$ gives us that

$$e_n = 2^n = \frac{2^{2n+1}}{2^{n+1}} = e_{n+1} \le f(x).$$

The last case is when $f(x) \le 2^n$. Then $e_n = 2^{-n}k_1$, and $e_{n+1} = 2^{-n-1}k_2$, where k_1, k_2 are both chosen so that $k_1 \in \{0, 1, \ldots, 2^{2n} - 1\}$, and $k_2 \in \{0, 1, \ldots, 2^{2n+2} - 1\}$. If $\frac{1}{2^{n+1}}(2k_1 + 1) < f(x)$, then we set $k_2 = 2k_1 + 1$, otherwise we let $k_2 = 2k_1$. This guarantees that $e_{n+1} \le f(x)$, however, in the former, $e_n < e_{n+1}$, while in the latter $e_n = e_{n+1}$. Now to show pointwise convergence. If $f(x) < +\infty$, then we have

$$e_n = 2^{-n}k \le f(x) < 2^{-n}(k+1) \tag{1}$$

$$\implies 0 \le f(x) - e_n < 2^{-n}. \tag{2}$$

Since the right hand side can be made arbitrarily small, we have convergence. Now if $f(x) = +\infty$, then $e_n(x) = 2^n$ for all n, giving us that $e_n \to +\infty$ as well, concluding the proof. ☺

Corollary 2.1. For *any* measurable function f, it is the pointwise limit of a sequence of simple functions. Further, if f is bounded, the convergence is uniform.

Proof. For a function f that is both negative and positive, we define $f_+(x) = f(x)$ whenever $f(x) \geq 0$ and 0 otherwise, similarly, we define $f_-(x) = -f(x)$ whenever $f(x) < 0$ and 0 otherwise. It is then easy to see that $f(x) = f_+ - f_-$, and that both f_+ and f_- are measurable. For $a \geq 0$, $f_+^{-1}((a,\infty]) = f^{-1}((a,\infty])$, a measurable set, and for $a < 0$, $f_+^{-1}((a,\infty]) = \{0\} \cup f^{-1}((a,\infty])$, also a measurable set. Since a difference of step functions is itself a step function, we can approximate f_+ and f_- separately, and thus also yield an approximation of f. For the last statement, let $\varepsilon > 0$. Then for some integer $n_0 \geq 1$, and for all x we have that $f(x) \leq 2^{n_0}$. Further, for all $n \geq n_0$, $e_n = 2^{-n}k$ for some k like above. Now for any such k, we have

$$\begin{aligned} & 2^{-n}k \leq f(x) \leq 2^{-n}(k+1) \\ \iff & 2^{-n}(k-1) \leq f(x) \leq 2^{-n}(k+1) \\ \iff & |f(x) - e_n(x)| \leq 2^{-n}, \end{aligned}$$

which gives us that the convergence is uniform. ☺

The nice thing about this theorem is that not only does it give us a positive statement about approximating measurable functions, it also gives us a way to compute these sequences. The way to visualize these constructed functions is to treat them as approximations of f up from *below*. In other words, $e_n \leq e_{n+1} \leq f$ for all n.

Step functions are pretty well-behaved, so if we can define integrals for them in a way that behaves nicely with limits we will essentially have made integrals for measurable functions well-behaved as well.

Definition 12. Assume that $(X, \mathcal{S}, \mu)$ is a measure space, and $e(x) = \sum_{i=1}^{n} \alpha_i \chi_{S_i}(x)$ a non-negative simple function on X. Then we define the integral of e on X with respect to the measure μ, as

$$\int_X e\, d\mu = \sum_{i=1}^{n} \alpha_i \mu(S_i).$$

We say that e is integrable if $\int_X e\, d\mu$ is finite.

If $E \subset X$ is a measurable set, we define the integral of e on E as

$$\int_E e\, d\mu = \int_X e\chi_E\, d\mu = \sum_{i=1}^{n} \alpha_i \mu(E \cap S_i).$$

One immediate consequence of this definition is that the integral on simple functions is a linear operation. Wanting to use this definition for defining a notion of integrability on measurable functions, we have the following.

Definition 13. Let $(X, \mathcal{S}, \mu)$ be a measure space and let $f : X \to [0,\infty]$ be a non-negative measurable function. The integral of f with respect to μ is defined as

$$\int_X f\, d\mu = \sup\left\{\int_X e\, d\mu : e \leq f,\, e \text{ simple}\right\}.$$

We call f integrable if $\int_X f\,d\mu < +\infty$ Further, if $E \subset X$ is a measurable set, we define the integral of f over E as

$$\int_E f\,d\mu = \int_X f\chi_E\,d\mu.$$

If f is both positive and negative, we define the integrable of f as

$$\int_X f_+\,d\mu - \int_X f_-\,d\mu,$$

and we say f is integrable if and only if both f_+ and f_- are finite. If instead, f is complex valued, then we say $f = g + ih$ is integrable if and only if g and h are integrable.

2.7 Elementary properties

For this section, whenever we mention a measurable function, it is implied it is part of a measure space $(X, \mathcal{S}, \mu)$. ' The following theorem tells us that Lebesgue integrals are order-preserving.

Theorem 3. Let $(X, \mathcal{S}, \mu)$ be a measure space, f, g be measurable functions from X to $[0, \infty]$, and E, F measurable sets. Then the following holds.

1. $0 \leq f \leq g$ implies $\int_E f\,d\mu \leq \int_E g\,d\mu$.
2. If $E \subset F$, then $\int_E f\,d\mu \leq \int_F f\,d\mu$.
3. If $E \cap F = \emptyset$, then $\int_{E\cup F} f\,d\mu = \int_E f\,d\mu + \int_F f\,d\mu$.

Proof.
1. By the definition of the supremum, $\sup(A) \leq \sup(B)$ for all $A \subset B$. For any simple function $e \leq f$, it also holds that $e \leq g$, and so the set of simple functions bounded above by f are a subset of the set of simple functions bounded above by g. Therefore,

$$\sup\left\{\int_X e\,d\mu : 0 \leq e \leq f\right\} \leq \sup\left\{\int_X e\,d\mu : 0 \leq e \leq g\right\}.$$

It follows by definition of the integral for non-negative functions that

$$\int_X f\,d\mu \leq \int_X g\,d\mu.$$

2. From the previous proof, note that $f\chi_F \leq f\chi_E$.
3. For simple functions $e = \sum_{k=1}^n c_k\chi_{A_k}$, we have that

$$\int_{E\cup F} \sum_{k=1}^n c_k\chi_{A_k}\,d\mu = \int_X \sum_{k=1}^n c_k\chi_{A_k}\chi_{E\cup F}\,d\mu.$$

Now it is an easy calculation to show that $\chi_{A_k}\chi_{E\cup F} = \chi_{A_k\cap E}\chi_{A_k\cap F}$ (since F and E are disjoint). In particular, this means that

$$\int_{E\cup F} \sum_{k=1}^n c_k\chi_{A_k}\,d\mu = \sum_{k=1}^n c_k\mu(A_k \cap E) + \sum_{k=1}^n c_k\mu(A_k \cap F) = \int_E e\,d\mu + \int_F e\,d\mu.$$

So the statement is true for simple functions. Now remember from Real Analysis that for $A, B \subset [0, \infty]$,

$$\sup\{a + b : a \in A, b \in B\} = \sup(A) + \sup(B),$$

and apply the result for

$$\begin{aligned}\int_{E\cup F} f\, d\mu &= \sup\left\{\int_{E\cup F} e\, d\mu : 0 \le e \le f\right\} \\ &= \sup\left\{\int_E e\, d\mu + \int_F e\, d\mu : 0 \le e \le f\right\} = \int_E f\, d\mu + \int_F f\, d\mu..\end{aligned}$$

And we are done.

☺

Another property we would want out of the Lebesgue integral is linearity. It appears that for measurable functions, this is quite tricky to prove without the use of heavier machinery. The next theorem will help us out, and is called the *Monotone Convergence Theorem*. The proof of this can be found in [AXL20, p. 78]. After having shown linearity, our previous results will also hold for non-negative functions, by taking $f = f_+ - f_-$.

Theorem 4 (Monotone Convergence)**.** Suppose that there is an increasing sequence f_n of measurable functions on X with values in $[0, \infty]$. Then the function f defined by the pointwise limit of f_n is nonnegative and measurable. Further, for all measurable sets E, we have

$$\lim_{n\to\infty} \int_E f_n\, d\mu = \int_E f\, d\mu.$$

Now we are ready to show that the Lebesgue integral is linear for non-negative functions.

Theorem 5. Suppose f, g are non-negative and measurable functions, a is any real number, and E any measurable set. Then

$$\int_E af\, d\mu = a\int_E f\, d\mu$$

and

$$\int_E f + g\, d\mu = \int_E f\, d\mu + \int_E g\, d\mu.$$

Proof. By definition, the Lebesgue integral is already linear for simple functions. Construct $\{e_n^{(f)}\}_{n\in\mathbb{N}}$ and $\{e_n^{(g)}\}_{n\in\mathbb{N}}$ as in the proof for Theorem 2. Then they are both increasing sequences of measurable functions, and their sum converging pointwise to $f + g$. By the Monotone Convergence theorem, we have that

$$\int_E f + g\, d\mu = \lim_{n\to\infty} \int_E e_n^{(f)} + e_n^{(g)}\, d\mu = \lim_{n\to\infty} \int_E e_n^{(f)}\, d\mu + \int_E e_n^{(g)}\, d\mu = \int_E f\, d\mu + \int_E g\, d\mu.$$

Similarly argued, we also get that for $a \ge 0$

$$\int_E af\, d\mu = \lim_{n\to\infty} \int_E ae_n^{(f)}\, d\mu = \lim_{n\to\infty} a\int_E e_n^{(f)}\, d\mu = a\int_E f\, d\mu.$$

If $a < 0$, f becomes non-positive, and so

$$\int_E -af\,d\mu = -\int_E (af)_-\,d\mu = -\int_E -af\,d\mu = a\int_E f\,d\mu$$

since $-a \geq 0$. ☺

Now linearity for arbitrary measurable functions follows by using the identity $f = f_+ - f_-$ and the linearity of non-negative functions. We also have that the triangle inequality holds for the Lebesgue integral.

Theorem 6. Let E be a measureable set, and f an integrable function, then

$$\left|\int_E f\,d\mu\right| \leq \int_E |f|\,d\mu.$$

Proof. The name hints at what to do:

$$\left|\int_E f\,d\mu\right| = \left|\int_E f_+\,d\mu - \int_E f_-\,d\mu\right| \leq \left|\int_E f_+\,d\mu\right| + \left|\int_E f_-\,d\mu\right| = \int_E |f|\,d\mu.$$

☺

Another consequence of the Monotone Convergence theorem, (that is usually used to show it) is called Fatou's Lemma. Reminder for the reader that

$$\liminf_{k\to\infty} x_k = \sup\{\inf\{x_j : j \geq k\} : k \geq 1\},$$

where $\inf\{x_j : j \geq k\}$ is an increasing sequence in k, and that the above is equivalent to

$$\liminf_{k\to\infty} x_k = \lim_{n\to\infty} \left(\inf\{x_k : k \geq n\}\right).$$

Theorem 7. Let $(X, \mathcal{S}, \mu)$ be a measure space and $\{f_k\}_{k\geq 1}$ is a sequence of non-negative measurable functions on X. Then the function $f : X \to [0, \infty]$ defined by $f(x) = \liminf_{k\to\infty} f_k(x)$ is measurable, and

$$\int_X f\,d\mu \leq \liminf_{k\to\infty} \int_X f_k\,d\mu.$$

Proof. We begin by noting that for each $n \geq 1$, $g_n(x) = \inf\{f_k(x) : k \geq n\}$ is a measurable function by Theorem 1, and so by the same theorem, $f(x) = \sup\{g_n(x) : n \geq 1\}$ is as well. By the definition of lim inf, this means that $f(x) = \liminf_{n\to\infty} f_k(x)$. Further, we have that

$$g_n(x) \leq f_k(x), \text{ for } k \geq n,$$

meaning there exists a non-decreasing subsequence $\{f_{k_n}\}_{n\geq 1}$ of f_k such that for each $x \in X$ and $n \geq 1$,

$$g_n(x) \leq f_{k_n}(x),$$

also implying that

$$\int_X g_n\,dx \leq \int_X f_{k_n}\,dx.$$

Since g_n is increasing, we can apply the Monotone Convergence theorem to yield that

$$\int_X f\,dx = \lim_{n\to\infty}\int_X g_n\,dx \le \lim_{n\to\infty}\int_X f_{k_n}\,dx.$$

But $\{\int_X f_{k_n}\,dx\}_{n\ge 1}$ is increasing as well! And so the limit in the right-hand side can be substituted with its $liminf$, concluding the proof. ☺

Theorem 8. Let $(X, \mathcal{S}, \mu)$ be a measure space, and suppose $f : X \to [-\infty, \infty]$ is measurable, then f is integrable if and only if $|f|$ is integrable.

Proof. By linearity of the integral, $|f| = f_+ + f_-$ is integrable if and only if both f_+ and f_- are integrable, which is equivalent to saying that f is integrable. ☺

For Riemann integrals, the above is not true in both directions. A good example is the function

$$f(x) = \chi_{[0,1]\cap\mathbb{Q}}(x) - \chi_{[0,1]\setminus\mathbb{Q}}(x).$$

Then $|f| = 1$ on $[0,1]$, and is therefore Riemann integrable, however, f clearly isn't as previous examples have shown.

Theorem 9. Let $(X, \mathcal{S}, \mu)$ be a measure space, and suppose $f : X \to [0, \infty]$ is nonnegative and measurable, then

$$\int_X f\,d\mu = 0 \text{ if and only if } f = 0 \text{ almost everywhere.}$$

What Theorem 9 is saying is that the Lebesgue integral does not discriminate between non-negative functions that only differ on a set of measure zero. From within it, any such functions are indistinguishable.

Now comes another very important result for the theory of the Lebesgue integral. It is used extensively throughout this text, and is called the *Dominated Convergence Theorem*.

Theorem 10. Let f_n be a sequence of measureable functions defined on X that converges almost everywhere to a function f. Suppose there exists an integrable function g such that for each $n \in \mathbb{N}$, $|f_n(x)| \leq g(x)$ a.e. on X. Then f is integrable, and further

$$\lim_{n\to\infty} \int_E f_n \, d\mu = \int_E f \, d\mu \quad \text{for all measurable sets } E \subset X.$$

Proof. [AXL20, p. 90] ☺

The Dominated convergence theorem gives us a more lenient condition for when we are able to pass a point-wise limit of a function inside an integral. Recall that for the Riemann integral, one often requires the limit to be uniform. When we integrate a function that is defined on a domain like $\mathbb{R} \times X$, where X is any measure space, we are essentially getting a new function of the form

$$I(t) = \int_X f(t,x)\, dx, t \in \mathbb{R}.$$

As we will see later this type of object comes up a lot when we start dealing with the Fourier transform on integrable functions, so we would like to know what kind of operations are permitted on these objects, and whether they have nice properties or not.

Theorem 11. Suppose that for almost all $x \in X$, $t \mapsto f(t,x)$ is continuous at a point t_0, and if there exists a an integrable function g such that for all t in a neighborhood U of t_0

$$|f(t,x)| \leq g(x) \text{ a.e.},$$

then $I(t) = \int_X f(t,x)\, dx$ is continuous at t_0.

Proof. Let $t_n \to t_0$ be a sequence in U, the sequence $f_n(x) = f(t_n, x)$ converges to $f(t_0, x)$ a.e. in X (since $h(t) = f(t,x)$ is continuous at t_0). Further, $|f_n(x)| \leq g(x)$ a.e., and so by Theorem 10, we have that

$$\lim_{n\to\infty} I(t_n) = \lim_{n\to\infty} \int_X f_n(x)\, d\mu = \int_X \lim_{n\to\infty} f_n(x)\, d\mu = I(t_0).$$

☺

The next theorem tells us when we can differentiate $I(t)$ under the integral sign.

Theorem 12. Suppose that U is a neighborhood of $t_0 \in \mathbb{R}$ such that

(i) $\frac{\partial f}{\partial t}$ exists almost everywhere on x, and is continuous on U.

(ii) There exists an integrable function g such that for all $t \in U$,

$$\left|\frac{\partial f}{\partial t}(t, x)\right| \leq g(x) \text{ a.e..}$$

Then I is differentiable at t_0, and $I'(t_0) = \int_X \frac{\partial f}{\partial t}\, d\mu$.

Proof. [GW99, p. 122] ☺

We end this section by stating *Fubini's Theorem*, because we would like to be able to interchange the order of integration sometimes.

Theorem 13. Let $f : \mathbb{R} \times \mathbb{R} \to [-\infty, \infty]$ be a measurable function. And $E \times F$ be a measurable set in $\mathbb{R} \times \mathbb{R}$.

1. If f is integrable on $E \times F$, then for almost every $x_0 \in E$, $f(x_0, y)$ is integrable on F, for almost every $y_0 \in F$, $f(x, y_0)$ is integrable on E, and the following holds

$$\int_{E\times F} f(x,y)\, dxdy = \int_E \left(\int_F f(x,y)\, dy\right) dx = \int_F \left(\int_E f(x,y)\, dx\right) dy.$$

2. f is integrable if and only if the integrals

$$\int_F \left(\int_E |f(x,y)|\, dx\right) dy \text{ or } \int_E \left(\int_F |f(x,y)|\, dy\right) dx$$

are finite.

Proof. [GW99, p. 124] ☺

There is still much to say on the Lebesgue integral, especially on how it relates to the Riemann integral, and things relating to principal values. We will leave this topic aside for the rest of this text and just trust that mathematicians did a good job at making sure that the whenever a function is Riemann-integrable, it is also Lebesgue integrable, and further, the integrals coincide. After all, what is the point of a new theory of integration if it doesn't at least encompass the older one?

3 Vinyls and Hats

3.1 L^p–spaces

Definition 14. Let $(X, \mathcal{S})$ be a measure space, and suppose $E \subset X$ is a measurable set. For $1 \le p < \infty$, we define $L^p(E)$ to be the class of measurable functions $f : E \to \mathbb{C}$ such that $|f|^p$ is integrable, in other words, so that

$$\int_E |f|^p \, d\mu < \infty.$$

Further, we define $L^\infty(E)$ as the class of measurable functions f with the property that for a constant $C \ge 0$ (dependent on f), $|f(x)| \le C$ a.e. on E, the elements of $L^\infty(E)$ are called essentially bounded.

The $L^p(E)$ spaces are vector spaces for all $1 \le p < \infty$. The only non-obvious property is to check they are closed under addition and scalar multiplication, but this follows from the linearity of the Lebesgue integral that we proved in the previous section. For $L^\infty(E)$, again recall that the union of two sets of measure 0 also has measure 0, and so by the triangle inequality

$$|f + g| \le |f| + |g| \le C + D \text{ a.e. on } E,$$

where $f, g \in L^\infty(E)$, and C and D are appropriate real numbers.

Defining a norm on $L^p(E)$ for $1 \le p < \infty$ is done by letting

$$\|f\|_p = \left(\int_E |f|^p \, d\mu \right)^{1/p}.$$

However, one should note that this is not actually a norm, since there are non-zero elements of $L^p(E)$ for which $\|f\|_p = 0$ (like the indicator function on the rationals we encountered earlier). However, we also know from Theorems 9 and 8, that if $\|f\|_p = 0$, then that means that $f = 0$ almost everywhere on E. Therefore, we can take a quotient with the subspace of all such functions, essentially identifying them with the zero function and end up with another vector space for which the above norm is well-defined. There are slight technicalities that have been ignored with this approach, but we march on. For $L^\infty(E)$, we define

$$\|f\|_\infty = \inf\{C \ge 0 : |f(x)| \le C \text{ a.e. on } E\}.$$

Theorem 14. Let $(X, \mathcal{S})$ be a measure space, and $U \subset X$ be a meaurable set such that $\mu(U) < \infty$. Then $L^2(U) \subset L^1(U)$

Proof. For any function f in $L^2(U)$ define the set $S = \{x \in U : |f| \ge 1\}$, then on S, $|f| \le |f|^2$, and on $U \setminus S$, $|f| \le \chi_U$ so

$$\begin{aligned} \int_U |f| \, d\mu &\le \int_S |f|^2 \, d\mu + \int_{U \setminus S} |f| \, d\mu \\ &\le \|f\|_2^2 + \mu(U) < \infty. \end{aligned}$$

Hence, $f \in L^1(U)$, and we are done. ☺

Definition 15. For some $a \in \mathbb{R}$, let $L_p^n(0,a)$ denote the space of a-periodic functions f such that $|f|^p$ is integrable on $(0,a)$,

Definition 16. Let $L^p_{\text{loc}}(\mathbb{R})$ denote the space of all measurable functions $f : \mathbb{R} \to \mathbb{R}$ such that

$$\int_a^b |f(x)|^p \, dx < +\infty \text{ for all } a < b.$$

It is clear (and the reader should confirm this), that for all $p \in [1,\infty)$, $L^p(\mathbb{R}) \subset L^p_{\text{loc}}(\mathbb{R})$.

Definition 17. Let $f : X \to \mathbb{C}$ be a continuous function. The *support* of f, denoted by $\text{supp}(f)$, is defined to be the closure of the set on which f is non-zero. In other words,

$$\text{supp}(f) = \overline{\{x \in X : f(x) \neq 0\}}.$$

Remark 3.

1. If f is defined on a subset of $\mathbb{R}$, and $\text{supp}(f)$ is a bounded set, then it also compact. We will often use this terminology in future sections.
2. Suppose f is a function with compact support, whose support is contained in some interval $(a, a+b)$. Then we can take the restriction of f on $(a, a+b)$, and extend the function $g(x) = f(x+a)$ into a b-periodic function, making it an element of $L_p^n(0,b)$ for any $n \in \{0,1,2,\ldots,\infty\}$

Definition 18. Let I be a subset of $\mathbb{R}$, and $p \in \{0,1,2,\ldots,\infty\}$. Define $C_c^p(I)$ as the set of functions that are differentiable up to order p with bounded support.

Example 19. The characteristic function on any bounded interval has compact support. More interesting examples are compactly supported smooth functions, also called bump functions. One such function is

$$f(x) = \begin{cases} e^{\frac{1}{|x|^2-1}} & \text{if } |x| < 1 \\ 0 & \text{if } |x| \geq 1 \end{cases}.$$

This function is infinitely differentiable on $(1,1)$.

Theorem 15. Let I be an open subset of $\mathbb{R}$. Then the spaces $C_c^0(I)$ are both dense in $L^1(I)$

Proof. [GW99, p. 138] ☺

3.2 Hilbert Spaces

Definition 20. Let V be a vector space over a field F(almost always $\mathbb{R}$ or $\mathbb{C}$). An operation $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{K}$ is called an inner product if it satisfies that for any $f, g, h \in V$ and $c \in F$

1. $\langle f, f \rangle \in [0, \infty)$,
2. $\langle f, f \rangle = 0$ if and only if $f = 0$,
3. $\langle cf + g, h \rangle = c \langle f, h \rangle + \langle g, h \rangle$,
4. $\langle f, g \rangle = \overline{\langle g, f \rangle}$.

A vector together with an inner product is called an *inner product space*.

Inner products are a good way to characterize the idea of *projecting* a vector onto another. From this, the idea of orthogonal vectors in an inner product space should also come to mind to the reader who has taken a course in linear algebra. We will almost exclusively work with the field $\mathbb{C}$ of complex numbers, so unless stated otherwise, F is always assumed to be $\mathbb{C}$.

Remark 4. Any inner product induces a norm by considering $\|f\| = \sqrt{\langle f, f \rangle}$. This means that any inner product space is also a normed space (and by extension, also a metric space).

The next two theorems are so important they get their own names. The first is called the Cauchy-Scwhartz inequality and gives a very useful bound on pairs of vectors in an inner product space. The second is called the Parallelogram identity.

Theorem 16. Let H be an inner product space. For any $f, g \in H$, we have that

1. $|\langle f, g \rangle| \leq \|f\| \|g\|$.
2. $\|f + g\|^2 + \|f - g\|^2 = 2(\|f\|^2 + \|g\|^2)$.

Example 21. Let $\ell^2(I)$ be the space of sequences with values in $\mathbb{C}$, indexed by some countable set I such that

$$\sum_{i \in I} |a_i|^2 < +\infty.$$

To see that this is indeed a vector space, note that for $\{a_i\}_{i \in I}, b = \{b_i\}_{i \in I} \in \ell^2(I)$, we have

$$\sum_{i \in I} |a_i + b_i|^2 \leq \sum_{i \in I} (|a_i| + |b_i|)^2 \leq 2 \sum_{i \in I} |a_i|^2 + |b_i|^2 < +\infty.$$

and so it is closed under addition. All the other properties are fairly obvious, and thus are omitted. What is not obvious, is that $\ell^2(I)$ is also an inner product space (and thus also a normed space). Define

$$\langle a, b \rangle_{\ell^2} = \sum_{i \in I} a_i \overline{b_i}. \tag{3}$$

We start by noting that $\langle \cdot, \cdot \rangle_{\ell^2}$ satisfies properties (1.–4.) pretty straightforwardly. However, it is not clear that the series in 3 even converges. For each finite subset K of I, we have that

$$\sum_{k \in K} \left| a_k \overline{b_k} \right| = \sum_{k \in K} \leq \sqrt{\sum_{k \in K} |a_k|^2} \sqrt{\sum_{k \in K} |b_k|^2},$$

by the Cauchy-Scwhartz inequality in $\mathbb{C}^{|K|}$. Taking limits as $K \to I$, the right-hand side is bounded, and so the series in 3 is absolutely convergent. For completion then, $(\ell^2(I), \langle \cdot, \cdot \rangle_{\ell^2})$ is an inner product space (and thus also a normed space).

Example 22. Let I be any interval of $\mathbb{R}$, then define for $f, g \in L^2(I)$, the product

$$\langle f, g \rangle_{L^2} = \int_I f(x)\overline{g}(x)\, dx.$$

Then it is not hard to show that this is indeed an inner product on $L^2(I)$. The fact that the standard norm on $L^2(I)$ is already induced by this should give us a hint. We start off by noting that if $|f\overline{g}| \in L^1(I)$, we are done. We have that

$$|f\overline{g}| = |fg| \leq \frac{|f|^2 + |g|^2}{2}.$$

Since the right hand side is in L^1, we have that $f\overline{g}$ also is. That $\langle f, f \rangle = 0 \iff f = 0$ in $L^2(I)$ has already been addressed earlier, and linearity and symmetry are fairly obvious.

If it is clear from the context, the symbol $\langle \cdot, \cdot \rangle$ to denote the inner product of a given inner product space.

Definition 23. An inner product space that is also a complete metric space with regards to its norm induced by the inner product is called a *Hilbert space*.

Theorem 17. The spaces $L^2(E)$ and $\ell^2(\mathbb{N})$ are both Hilbert spaces.

Proof. ☺

Definition 24. Let H be an inner product space, and $x, y \in H$.

1. We call x and y orthogonal, denoted by $x \perp y$, if $\langle x, y \rangle = 0$.
2. For any two subsets U, W of a hilbert space H, they are called orthogonal, denoted by $U \perp W$, if every element in U is orthogonal to all elements of W.
3. For a subset $U \subset H$, we define the orthogonal complement of S as

$$S^{\perp} = \{y \in H : \langle x, y \rangle = 0 \text{ for all } x \in S\}.$$

The following identity is called Pythagoras' identity for inner product spaces.

Theorem 18. Let H be an inner product space, and $x_1, x_2, \ldots, x_n \in H$. If $x_i \perp x_j$ for $i \neq j$, then

$$\left\| \sum_{k=1}^{n} x_k \right\|^2 = \sum_{k=1}^{n} \|x_k\|^2.$$

Proof. We cover the case for $n = 2$, and then let $x_2 + \ldots + x_n$ be a single vector and be done. Starting with the left hand side, we have that

$$\begin{aligned} \|x + y\|^2 =& \langle x, x + y \rangle + \langle y, x + y \rangle \\ =& \|x\|^2 + \langle x, y \rangle + \overline{\langle x, y \rangle} \|y\|^2 \\ =& \|x\|^2 + \|y\|^2. \end{aligned}$$

☺

The next theorem let's us characterize the element for a given vector x that minimizes the distance between them. This also captures our well-known intuition that such an element should be orthogonal to x.

Theorem 19. Let H be a Hilbert space $x \in H$, and M a closed subspace of H, then there exists a unique non-zero element $y \in M$ such that

$$\|x - y\| = \min_{p \in M} \|x - p\|.$$

Further, y is the unique (non-zero) element of M such that $\{x - y\} \perp M$.

Proof. [GW99, p. 143] ☺

The next theorem is something statisticians will recognize all too well.

Theorem 20. Suppose H is a hilbert space, V is a a subspace of H, and $f \in H$. Then an element $f^* \in V$ satisfies

$$\|f - f^*\| = \min_{v \in V} \|f - v\| \quad \text{if and only if} \quad \langle f - f^*, w \rangle = 0 \text{ for all } w \in V.$$

Proof. [GW99, p. 144] ☺

We call $c_n(f) = \langle f, \varphi_n \rangle / \|\varphi_n\|^2$ the Fourier coefficient of f relative to φ_n.

We would like to have a criteria for when a series of vectors (viewed as the limit of a sequence in a Hilbert space), actually converges to a vector. We have a Cauchy criterion, (since any Hilbert space is complete), but here is a nice result one can make use of as long as the vectors in the series are orthonormal.

Lemma 21. Let H be a Hilbert space, $\{\varphi_n\}_{n=1}^{\infty}$ a sequence of pairwise-orthonormal vectors in H, and $\{a_n\}_{n=1}^{\infty} \in \mathbb{C}^{\mathbb{N}}$. Then any series of the form

$$\sum_{n=1}^{\infty} a_n \varphi_n$$

converges in H if and only if

$$\sum_{n=1}^{\infty} |a_n|^2 < \infty.$$

Proof. It follows from the completeness of Hilbert spaces that for sufficiently large l, k with $l > k$, our sequence converges if and only if

$$\left\| \sum_{n=k}^{l} a_n \varphi_n \right\|^2 \text{ approaches } 0.$$

By Pythagoras' (Theorem 18), since φ_n are all pairwise orthonormal, we get

$$\left\| \sum_{n=k}^{l} a_n \varphi_n \right\|^2 = \sum_{n=k}^{l} |a_n|^2 .$$

It follows then that the moment the right hand side approaches zero, so must the left hand side, so our statement is proven. ☺

Theorem 22 (Bessel's Inequality)**.** Suppose H is Hilbert space and that for a sequence $\{\varphi_n\}_{n\in\mathbb{N}}$ of pairwise-orthogonal elements in H then for all $f \in H$,

$$\sum_{n=1}^{\infty} |c_n(f)|^2 \|\varphi_n\|^2 \leq \|f\|^2,$$

where $c_n(f)$ are all Fourier coefficients.

Proof. Let $f_p = \sum^{p} c_n(f)\varphi_n$. Note that since all elements in $\Phi = \{\varphi_1, \ldots, \varphi_p\}$ are pairwise orthogonal, they are also linearly independent, and so they form a basis. Further, it follows by linearity in the first argument that $\langle f - f_p, f_p \rangle = 0$, and so by Theorem 20, we have that f_p is the projection of f onto the linear span of Φ. Expanding $\|f - f_p\|^2$, we get

$$\begin{aligned} \langle f - f_p, f - f_p \rangle &= \|f\|^2 - \langle f, f_p \rangle \\ &= \|f\|^2 - \left\langle f, \sum_{n=1}^{p} c_n(f)\varphi_n \right\rangle \\ &= \|f\|^2 - \sum_{n=1}^{p} \overline{c_n(f)} \langle f, \varphi_n \rangle . \end{aligned}$$

Notice that $\langle f, \varphi_n \rangle = c_n(f) \cdot \|\varphi_n\|^2$, so we finally get that

$$0 \leq \|f - f_p\|^2 = \|f\|^2 - \sum_{n=1}^{p} |c_n(f)|^2 \|\varphi_n\|^2.$$

Letting $p \to \infty$, we get our result, concluding our proof. ☺

What Bessel's inequality is telling us, is that for any $f \in H$, if our coefficients c_n in Lemma 21 are of the form $\langle f, \varphi_n \rangle$, where φ_n denotes an element of an indexed set of pairwise-orthogonal elements, then we know for sure that our series converges. We still haven't established what they converge *to*, but we might have a hunch. . .

Definition 25. We call a sequence $\{\varphi_n\}_{n=1}^{\infty}$ of pairwise-orthonormal elements of a Hilbert space H an *orthonormal basis* if for every $f \in H$, the following holds

$$f = \sum_{n=1}^{\infty} c_n(f)\varphi_n.$$

Next theorem is called Parseval's identity and will be used in the upcoming sections.

Theorem 23 (Parseval's identity)**.** Let $\{\varphi_n\}_{n=1}^{\infty}$ be an orthonormal basis and H a Hilbert space, then for every $f \in H$,

$$\|f\|^2 = \sum_{n=1}^{\infty} |c_n(f)|^2.$$

Proof. For any integer p, let f_p denote the projection of f onto the subspace spanned by $\{\varphi_1, \ldots, \varphi_p\}$ as before. Then we already know that $f - f_p$ is orthogonal to f_p, and so

$$\|f\|^2 = \|f - f_p + f_p\|^2 = \|f - f_p\| + \|f_p\|^2.$$

Moving a term, and using Pythagoras again, we get

$$\|f\|^2 - \sum_{n=1}^{p} |c_n|^2 = \|f - f_p\|^2.$$

Letting $p \to \infty$, the right hand side goes to 0, and we get our desired result. ☺

What about the other way around? If we have an orthonormal sequence of vectors $\Phi = \{\varphi_n\}_{n=1}^{\infty}$, and for every $f \in H$, $\|f\|^2 = \sum_{n=1}^{\infty} |c_n(f)|^2$, is it true that

$$\sum_{n=1}^{\infty} c_n(f)\varphi_n = f \quad \text{in } L^2(\mathbb{R})? \tag{4}$$

It turns out that the answer is yes! First we note that only the zero vector can have the property that $\langle f, \varphi_n \rangle = 0$ for all $\varphi_n \in \Phi$. Therefore, let φ_N be an arbitrary element of Φ, then

$$\left\langle f - \sum_{n=1}^{\infty} c_n(f)\varphi_n, \varphi_N \right\rangle = \langle f, \varphi_N \rangle - \sum_{n=1}^{\infty} c_n(f) \langle \varphi_n, \varphi_N \rangle = \langle f, \varphi_N \rangle - c_N(f).$$

In this case, $c_N(f)$ is just $\langle f, \varphi_N \rangle$, since $\|\varphi_N\| = 1$. The equality given in (4) follows.

In the upcoming sections, we will see multiple examples of orthonormal bases of Hilbert spaces.

4 Enough Fourier analysis to get by

We can use the information we have developed about Hilbert spaces in the previous sections to discuss the convergence of sequences of functions that make up an orthonormal basis for $L^2_p(0, 2\pi)$, the space of all square-integrable complex-valued functions with period 2π, with the inner product defined by

$$\langle f, g \rangle = \int_0^{2\pi} f(t)\overline{g}(t)\, dt,$$

as well as the norm induced by it:

$$\|f\|_2 = \sqrt{\langle f, f \rangle} = \left(\int_0^{2\pi} |f(t)|^2 \, dt \right)^{1/2},$$

after identifying all functions that are identically 0 a.e. with the zero functions (as otherwise the above would not be a norm). Then we have from earlier that $L^2_p(0, 2\pi)$ is a Hilbert space, and a nice orthogonal basis is the set of functions $\{e^{int}\}_{n\in\mathbb{Z}}$, as we shall now see.

Theorem 24. The set of trigonometric polynomials $\{e^{ikt}\}_{k\in\mathbb{Z}}$ is an orthogonal basis for $L^2_p(0, 2\pi)$.

Proof. [GW99, p. 150] ☺

4.1 The Fourier Transform on $L^1(\mathbb{R})$.

Definition 26. Let $f \in L^1(\mathbb{R})$, then we define the continuous linear transformations

$$\mathscr{F}(f) = \widehat{f}(\xi) = \int_{\mathbb{R}} e^{-i\xi x} f(x)\, dx, \tag{5}$$

$$\overline{\mathscr{F}} f(\xi) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\xi x} f(x)\, dx. \tag{6}$$

We call (5) the Fourier Transform of f and (6) the Conjugate Fourier transform (note that $\overline{e^{iy}} = e^{-iy}$).

Example 27. The Fourier transform of $\chi_{[a,b]}$ is given by

$$\begin{aligned}
\mathscr{F}(\chi_{[a,b]}) &= \int_a^b e^{ix\xi}\, dx \\
&= \frac{e^{-ia\xi} - e^{-ib\xi}}{i\xi} \\
&= e^{-i(a+b)\xi/2} \frac{e^{i(b-a)\xi/2} - e^{-i(b-a)\xi/2}}{i\xi} \\
&= e^{-i\xi\frac{a+b}{2}} \frac{\sin(\xi(b-a)/2)}{\xi/2}, \text{ for } \xi \in \mathbb{R} \setminus \{0\},
\end{aligned}$$

or if one prefers, the more compact form,

$$(b-a)e^{-i\xi\frac{a+b}{2}} \operatorname{sinc}(\frac{\xi}{2}(b-a)). \tag{7}$$

Theorem 25 (Riemann-Lebesgue theorem)**.** If $f \in L^1(\mathbb{R})$, then $\widehat{f}(\xi)$ satisfies that

(i) $\mathscr{F}$ is a contintuous linear operator from $L^1(\mathbb{R})$ to $L^\infty(\mathbb{R})$, and

$$\|\widehat{f}\|_\infty \leq \|f\|_1.$$

(ii) $\lim_{|\xi|\to\infty} \left|\widehat{f}(\xi)\right| = 0.$

Proof. (i) partly follows from Theorem 11, saying that the lebesgue integral is a continuous operation. That, together with the map $\xi \mapsto e^{-i\xi x} f(x)$ also being continuous, and

$$\left|\widehat{f}(\xi)\right| = \left|\int_{-\infty}^{\infty} e^{-i\xi x} f(x)\, dx\right| \leq \int_{-\infty}^{\infty} |f(x)|\, dx = \|f\|_1 < +\infty,$$

means that $\widehat{f} \in L^\infty(\mathbb{R})$. To show (ii), first we note from earlier, that the simple functions are dense in $L^1(\mathbb{R})$, (Theorem 2), and for each such function, it is clear by our previous example, that $\lim_{|\xi|\to\infty} \widehat{e}(\xi) = 0$, for any simple function $e(x)$. Therefore, for $f \in L^1(\mathbb{R})$, let e_n be a sequence of simple functions such that $\lim_{n\to\infty} \|f - e_n\|_1 = 0$. From our proof in (i), we have that

$$\left|\widehat{f} - \widehat{e_n}\right| \leq \|f - e_n\|_1$$

for each ξ, meaning that as $|\xi| \to \infty$, and for large enough n, we can make $|\widehat{f}(\xi)|$ as small as we desire, which concludes the proof. ☺

An important result that will be used to discuss the Inverse of the Fourier Transform is the following.

Theorem 26. Let f and g be in $L^1(\mathbb{R})$. Then $f\widehat{g}$ and $\widehat{f}g$ are both in $L^1(\mathbb{R})$. Further we have that

$$\int f(t)\widehat{g}(t)\,dt = \int \widehat{f}(x)g(x)\,dx.$$

Proof. As we saw in Theorem 25, $\widehat{g}$ is bounded, and so $f\widehat{g}$ will be integrable, and the same applies to $\widehat{f}g$. The equality follows from a simple application of Fubini's theorem (Theorem 13). ☺

It is not hard to see that of the properties discussed for $\mathscr{F}$, the same hold for $\overline{\mathscr{F}}$

4.2 Rules for computing with the Fourier transform

For the Fourier transform, derivatives play somewhat nicely, as showcased in the following theorem.

Theorem 27.

(i) If $x^k f(x)$ are in $L^1(\mathbb{R})$, for $k = 0, 1, 2, \ldots, n$, then $\widehat{f}$ is n times differentiable, and its derivatives are decided by

$$\widehat{f}^{(k)} = \mathscr{F}\left((-ix)^k f(x)\right).$$

(ii) If $f \in C^n(\mathbb{R}) \cap L^1(\mathbb{R})$, and $f^{(k)} \in L^1(\mathbb{R})$ for all $k = 1, 2, \ldots, n$, then

$$\widehat{f^{(k)}}(\xi) = (i\xi)^k \widehat{f}(\xi) \text{ for } k = 1, 2, \ldots, n.$$

(iii) If $f \in L^1(\mathbb{R}) \cap C_c^0(\mathbb{R})$, then $\widehat{f} \in C^\infty(\mathbb{R})$,

The last point tells us that the Fourier transform of an integrable function with bounded support is actually analytic.)

Proof. [GW99, p. 157] ☺

Example 28. Let $f \in L^1(\mathbb{R})$ and $a \in \mathbb{R}$, then the following holds:

1. $\mathscr{F}(f(x-a))(\xi) = e^{-ia\xi}\mathscr{F}(f(x))(\xi)$.
2. $\mathscr{F}(f(x))(\xi - a) = \mathscr{F}(e^{iax}f(x))(\xi)$.
3. If $a \neq 0$ $\mathscr{F}(f(ax))(\xi) = \frac{1}{|a|}\mathscr{F}(f(x))(\xi/a)$

Proof. For all three it is just a matter of change of variables

$$\mathscr{F}(f(x-a))(\xi) \int_{-\infty}^{\infty} e^{-ix\xi} f(x-a)\,dx = \int_{-\infty}^{\infty} e^{-i(y+a)\xi} f(y)\,dy = e^{-ia\xi}\mathscr{F}(f(y))(\xi).$$

Similarly,

$$\mathscr{F}(f(x))(\xi - a) = \int_{-\infty}^{\infty} e^{-ix(\xi-a)} f(x)\, dx = \int_{-\infty}^{\infty} e^{-ix\xi} \left(e^{iax} f(x)\right) dx = \mathscr{F}\left(e^{iax} f(x)\right).$$

Now suppose $a \neq 0$, then we have

$$\mathscr{F}(f(ax))(\xi) = \int_{-\infty}^{\infty} e^{-ix\xi} f(ax)\, dx = \int_{-\infty}^{\infty} e^{-iy\xi/a} f(y) \frac{dy}{|a|} = \frac{1}{|a|}\mathscr{F}(f(y))(\xi/a).$$

☺

If the reader is asking why the author seems undecided between using $\mathscr{F}(f)$ and $\widehat{f}$, it is answered by what looks more readable at a given moment, for the most part we will stick to using $\widehat{f}$ from now.

4.3 Inverse Fourier Transform

Theorem 28. If both f and $\widehat{f}$ are integrable (that is, in $L^1(\mathbb{R})$, then at each point where f is continuous,

$$f(x) = \overline{\mathscr{F}}\left(\widehat{f}(\xi)\right)(x).$$

Proof. [GW99, p. 163] ☺

4.4 The Schwartz Class of functions

Here we develop some of the theory behind an interesting class of functions that lies in $L^p(\mathbb{R})$, for all $p \in [1, \infty]$, namely the *Schwartz class*. These are functions that lie in $C^\infty(\mathbb{R})$, and that are particularly well behaved with regards to integration, as we shall see. This space is also dense in all $L^p(\mathbb{R})$, $p \in [1, \infty)$ (see [GW99, p. 189]) making it a good bridge for extending operators like the Fourier transform on other spaces other than $L^1(\mathbb{R})$. Let us therefore define what it means for a function to decay rapidly.

Definition 29. If a function $f : \mathbb{R} \to \mathbb{C}$ has the property that

$$\lim_{|x|\to\infty} |x^n f(x)| = 0 \text{ for all } n = 0, 1, 2, \ldots,$$

we say that f decays rapidly.

From our earlier discussion on the Fourier transform, and how it interacts with derivatives, we can say a few things about functions that are rapidly decaying.

Lemma 29. If $|f|$ is locally integrable and rapidly decaying, then $x^n f(x)$ is in $L^1(\mathbb{R})$ for all $n \in \mathbb{N}$.

Proof. Let M be such that $\left|x^{n+2}f(x)\right| \leq 1$ whenever $|x| > M$. Then we have that

$$\int_{\mathbb{R}} |x^n f(x)|\, dx \leq \int_{|x|\leq M} |x^n f(x)|\, dx \int_{|x|>M} \frac{1}{x^2} \left|x^{n+2}f(x)\right|\, dx$$
$$\leq M^n \int_{|x|\leq M} |f(x)|\, dx + \int_{|x|>M} \frac{1}{x^2}\, dx < +\infty.$$

☺

Theorem 30. Suppose f is integrable and decays rapidly, then $\widehat{f}$ is infinitely differentiable.

Proof. Since f is integrable, it is also locally integrable, and so $x^n f$, for $n \in \mathbb{N}$ is also in $L^1(\mathbb{R})$ by our previous lemma. It follows then from Theorem 27, that $\widehat{f} \in C^\infty(\mathbb{R})$ (since the Fourier transform is also continuous). ☺

Corollary 30.1. If $f \in C^\infty(\mathbb{R})$, then $\widehat{f}$ decays rapidly.

Proof. This follows from the Riemann-Lebesgue theorem, and knowing that $\widehat{f^{(k)}}(\xi) = (2\pi\xi)^k \widehat{f}(\xi)$. In particular, we get that if f is both in $C^\infty(\mathbb{R})$ and decays rapidly, then the same holds for $\widehat{f}$. ☺

If only there was a space for which the above property would turn the Fourier transform into a bijection. . . .

Definition 30. Let $\mathscr{S}$ denote the vector space of functions $f : \mathbb{R} \to \mathbb{C}$ in $C^\infty(\mathbb{R})$ with rapidly decaying derivatives (as well as f itself being rapidly decaying). We will also call $\mathscr{S}$ the Schwartz class of functions.

Theorem 31. The space $\mathscr{S}$ is

(i) closed under differentiation,

(ii) closed under multiplication by a polynomial,

(iii) contained in $L^1(\mathbb{R})$.

(iv) closed under Fourier transformations, that is, $f \in \mathscr{S} \implies \widehat{f} \in \mathscr{S}$.

Proof. For (i), we notice that the derivative of a function $f \in \mathscr{S}$ is also rapidly decaying, as well as infinitely differentiable, so $f' \in \mathscr{S}$. For $p(x) = \sum_{k=0}^{n} a_k x^k$,

$$|f(x)p(x)| \leq \sum_{k=0}^{n} \left|a_k x^k f(x)\right|.$$

Showing (ii), it is well-known that taking a limit is a linear operation, so letting $|x| \to \infty$ we get that $p(x)f(x) \in \mathscr{S}$. The proof of (iii) follows from the fact that any differentiable function is automatically locally integrable and then using Theorem 29. Lastly, suppose $f \in \mathscr{S}$, then we have shown that $\widehat{f} \in C^{\infty}$ (Theorem 30). Since $f^{(k)}$ is in $\mathscr{S}$ by (i), for all $k \in \mathbb{N}$, we get from Corollary 30.1 that $\widehat{f}$ decays rapidly. What we need now is to show that all derivatives of $\widehat{f}$ also decay rapidly, that is, $\lim\limits_{|\xi|\to\infty} \left|\xi^p \widehat{f}^{(n)}(\xi)\right| = 0$. We have by Theorem 27 that

$$\xi^p \widehat{f}^{(n)}(\xi) = \xi^p \mathscr{F}\left((-ix)^n f(x)\right)(\xi). \tag{8}$$

By the same theorem know that the right hand side of (8) equals

$$\frac{1}{(i)^p}\mathscr{F}\left(\left((-ix)^n f(x)\right)^{(p)}\right),$$

and by Theorem 25 this all goes to 0, which is precisely what we wanted to show. ☺

Remark 5. If a function is infinitely differenitable and has compact support, then it is in $\mathscr{S}$. If a function f has compact support, so will its derivatives, as well as any new function arising via multiplication of f by a polynomial.

4.5 Inverse Fourier Transform on $\mathscr{S}$

Thanks to Theorem 28 we know that for all functions f in $\mathscr{S}$, and for all $x \in \mathbb{R}$, $f = \overline{\mathscr{F}}\left(\mathscr{F} f\right)$. But then $\mathscr{F}\left(\overline{\mathscr{F}} f\right) = f$ as well, and so we see that the Fourier transform is a bijection on $\mathscr{S}$.

4.6 The Fourier Transform on $L^2(\mathbb{R})$

It is a well-known fact (if not, then read [AXL20, p. 169]) that an operator T between two normed spaces X and Y is continuous, if and only if for all $x \in X$,

$$\|Tx\|_Y \leq k\|x\|_X, \text{ for some } k \in \mathbb{R}.$$

We begin with a fun theorem about operators of complete normed spaces.

Theorem 32. Let X be a complete, normed space, and $T : \mathcal{S} \to X$ a linear operator, where $\mathcal{S}$ is a dense subspace of X, satisfying for all $x \in \mathcal{S}$ that $\|Tx\| \leq k\|x\|$ for some constant $k > 0$. Then there exists a linear operator $\widetilde{T} : X \to X$ such that the restriction of $\widetilde{T}$ to $\mathcal{S}$ equals T, and such that for all $x \in X$, $\|\widetilde{T}x\| \leq k\|x\|$.

Proof. Define $\widetilde{T}x = Tx$ for $x \in \mathcal{S}$, and let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{S}$ converging to x. Then Tx_n is Cauchy by the following argument: Let $\varepsilon > 0$, and choose n_0 such that for all $n, m > n_0$, $\|x_n - x_m\| < \frac{\varepsilon}{k}$. Then $\|Tx_n - Tx_m\| = \|T(x_n - x_m)\| \leq \varepsilon$, by the continuity of T. So $\{Tx_n\}_n$ converges in X, so call this limit $\widetilde{T}x$. To show that this limit is independent of the chosen sequence, let $x_n \to x$ be as before, and now consider another sequence in $\mathcal{S}$, $y_n \to x$, and let $\widetilde{T}y = \lim\limits_{n\to\infty} Ty_n$. Then $\widetilde{T}y = \widetilde{T}x$. Again, let $\varepsilon > 0$, and now choose n_0 such that for all $n > n_0$, we have

$$\|y_n - x\| < \frac{\varepsilon}{3k}, \ \|x_n - x\| < \frac{\varepsilon}{3k}, \text{ and } \|Tx_n - \widetilde{T}x\| < \frac{\varepsilon}{3}.$$

Then

$$\begin{aligned}\|Ty_n - \widetilde{T}x\| \leq & \|Ty_n - Tx_n\| + \|Tx_n - \widetilde{T}x\| \\ < & k\|y_n - x_n\| + \frac{\varepsilon}{3} \\ < & k\left(\|y_n - x\| + \|x - x_n\|\right) + \frac{\varepsilon}{3} < \varepsilon.\end{aligned}$$

Now we want to show that $\|\widetilde{T}x\| \leq k\|x\|$. *Again*, let $\varepsilon > 0$, and choose n_0 so that $\|\widetilde{T}x - Tx_n\| < \varepsilon$. Then we have

$$\|\widetilde{T}x\| \leq \|\widetilde{T}x - Tx\| + \|Tx_n\| < \varepsilon + k\|x_n\|.$$

Letting $n \to \infty$ yields us our inequality. Finally, linearity just follows from the fact that taking limits is a linear operation, concluding the proof. ☺

Corollary 32.1. The Fourier Transform can be extended to any function of $L^2(\mathbb{R})$.

This follows by applying Theorem 32 with $S = \mathscr{S}$ and $X = L^2(\mathbb{R})$. From now on, $\mathscr{F}(f)$ and $\widehat{f}$ will always refer to the extended Fourier transform on $L^2(\mathbb{R})$. This extension inherits with it a lot of the properties discussed in the last section on $\mathscr{S}$.

Theorem 33. The Fourier transform extended to $L^2(\mathbb{R})$ satisfies the following properties for all f and g in $L^2(\mathbb{R})$.

1. $\mathscr{F}\overline{\mathscr{F}}f = \overline{\mathscr{F}}\mathscr{F}f = f$ a.e.
2. $\int_{\mathbb{R}} f\overline{g}\, dx = \int_{\mathbb{R}} \mathscr{F}(f)\overline{\mathscr{F}(g)}\, d\xi$.
3. $\|f\|_2 = \frac{1}{2\pi}\|\mathscr{F}f\|_2$.
4. $\mathscr{F}(f \cdot g), f \cdot \mathscr{F}(g) \in L^1(\mathbb{R})$, and

$$\int_{\mathbb{R}} \mathscr{F}(f \cdot g)\, dx = \int_{\mathbb{R}} f \cdot \mathscr{F}(g)\, dx.$$

The third statement is called Plancherel's theorem, and will be used extensively throughout the upcoming section.

Proof. [GW99, p. 194] ☺

5 Wavelets, or, we are just getting started!

5.1 Introduction to Wavelets

Let $\varphi \in L^2(\mathbb{R})$. Throughout this section, we will be talking about doubly indexed sequences of the kind

$$\varphi_{j,k}(x) = 2^{j/2}\varphi\left(2^j x - k\right), \; j,k \in \mathbb{Z}.$$

Here we are capturing the idea of shifting φ around by changing k, and dilating it by changing j. The reason we want to multiply by a factor of $2^{j/2}$ is to preserve the L^2–norm with the original function. That is, we get that

$$\|\varphi_{j,k}\|_2 = \|\varphi\|_2$$

Derived from a combination of identities of the Fourier transform, we will be using the following identity a lot during this section for $\varphi \in L^2(\mathbb{R})$

$$\widehat{\varphi}_{j,k}(\xi) = 2^{-j/2} e^{-i2^{-j}\xi k} \widehat{\varphi}(2^{-j}\xi) \tag{9}$$

This equality is not too difficult to derive, we first assume $\varphi_{j,k}(x) \in \mathscr{S}$. Following the symbols, we get

$$\begin{aligned}\widehat{\varphi}_{j,k}(\xi) &= \int_{\mathbb{R}} e^{-i\xi x} 2^{j/2} \varphi\left(2^j x - k\right)\, dx \\ &= \int_{\mathbb{R}} e^{-i\xi(y+k)2^{-j}} 2^{-j/2} \varphi(y)\, dy \\ &= 2^{-j/2} e^{-i2^{-j}\xi k} \widehat{\varphi}(2^{-j}\xi).\end{aligned}$$

Therefore the result holds in $\mathscr{S}$. Now suppose $\varphi \in L^2(\mathbb{R})$, and $\{f_n\}_{n\geq 1}$ a sequence in $\mathscr{S}$ converging to φ in $L^2(\mathbb{R})$. Then for each $j, k \in \mathbb{Z}$, $\{(f_n)_{j,k}\}_{n\geq 1}$ converges to $\varphi_{j,k}$ in $L^2(\mathbb{R})$ as well. Note also that by Plancherel's Theorem (Theorem 33.3), the Fourier transform of $\{(f_n)_{j,k}\}_{n\geq 0}$ also converges to $\widehat{\varphi_{j,k}}$ in $L^2(\mathbb{R})$. Therefore, let N be large enough so that for $n \geq N$,

$$\|f_n - \varphi\|_2 < \varepsilon \frac{2^{j-1}}{\pi}.$$

Then for the same $n \geq N$, we have

$$\begin{aligned}\|\widehat{(f_n)_{j,k}} - 2^{-j/2} e^{-i2^{-j}\xi k} \widehat{\varphi}\|_2 &= 2^{-j} \|e^{-i2^{-j}\xi k} (\widehat{f}_n - \widehat{\varphi})\|_2 \\ &= 2^{-j+1} \pi \|f_n - \varphi\|_2 < \varepsilon.\end{aligned}$$

Definition 31. Let $\varphi \in L^2(\mathbb{R})$, if the sequence $\{\varphi_{j,k}\}_{j,k\in\mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{R})$, we call φ a *wavelet*, and the sequence $\{\varphi_{j,k}\}_{j,k\in\mathbb{Z}}$ a *wavelet basis*.

Before giving an example of a wavelet, a motivation for Wavelets is that $L^2(\mathbb{R})$ does not have many nicely behaved bases. For periodic, square-integrable functions, the elements $\{e^{inx}\}_{n\in\mathbb{Z}}$ suffice to have an ON-basis. However, this idea cannot be extended to the non-periodic case. As a matter of fact, e^{ixn} is not even in $L^2(\mathbb{R})$!

From what we have developed about Hilbert spaces, and in particular, $L^2(\mathbb{R})$, we can treat Wavelets as more sophisticated versions of our well-known, trigonometric polynomials. That is, we can project a given function f in $L^2(\mathbb{R})$ in a way that actually makes sense. Further, we will also be able to talk about projecting functions onto subspaces of $L^2(\mathbb{R})$, spanned by subsets of our Wavelet basis, another thing that doesn't make sense when making use of trigonometric polynomials, as again, $e^{-ix\xi}$ is not in $L^2(\mathbb{R})$.

Example 32 (Haar wavelet)**.** Consider the function

$$\psi(x) = \begin{cases} 1 & \text{if } x \in [0, \frac{1}{2}] \\ -1 & \text{if } x \in [-\frac{1}{2}, 0) \end{cases},$$

This compactly supported step function is called the Haar function, and has the property that

$$\int_{-\infty}^{\infty} \psi_{n,k}(x)\, dx = 0 \text{ for all } n, k \in \mathbb{Z}.$$

Further, for any fixed j, and $k, k' \in \mathbb{Z}$, $\psi_{j,k}$ and $\psi_{j,k'}$ never overlap in support, meaning that

$$\left\langle \psi_{j,k}, \psi_{j,k'} \right\rangle = \begin{cases} 0 & \text{if } k \neq k' \\ 1 & \text{if } k = k' \end{cases}.$$

We will see by the end of this section that this orthonormal sequence actually forms a basis of $L^2(\mathbb{R})$. To get there, we will first discuss how to divide $L^2(\mathbb{R})$ into different approximation spaces with some desirable properties.

Definition 33 (Multiresolution analysis)**.** Suppose $\{V_j\}_{j\in\mathbb{Z}}$ is a sequence of closed subspaces of $L^2(\mathbb{R})$, such that

(i) Every V_j is contained in V_{j+1}, for all $j \in \mathbb{Z}$;

(ii) the countable union $\bigcup_{j\in\mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R})$;

(iii) $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$;

(iv) given a function f in V_j, $g = f(2^{-j})$ is in V_0 and vice-versa;

(v) if $f \in V_0$, then $f_{0,k} \in V_0$ for all $k \in \mathbb{Z}$;

(vi) there exists a function $\varphi \in V_0$ such that the sequence $\{\varphi_{0,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis for V_0.

Then the sequence $\{V_j\}_{j\in\mathbb{Z}}$ is called a *Multiresolution analysis*, and we call each element V_j an *approximation space*.

A way to view a Multiresolution analysis is, as the definition suggests, by considering each V_j in the sequence as a finer and finer approximation of a function in $L^2(\mathbb{R})$ the larger j gets (and of course, coarser the smaller j gets). Viewed through this lens, axiom (i) tells us coarser resolution should be contained in finer ones. Axiom (ii) tells us that together, these resolutions make up the whole space, or at least a dense subset of it, Axiom (iii) gives us a condition on the resolutions losing the ability to distinguish from the trivial function the coarser you get. Axioms (iv) and (vi) gives us a way to jump between resolutions via scaling, as it forces each V_j to contain a scaled version of each function in V_0, for which we have a basis. Note that this means that for each j, $\{\varphi_{j,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis for V_j. Orthonormality is given by a simple change of variables, let $k, k' \in \mathbb{Z}$, then

$$\left\langle \varphi_{j,k}, \varphi_{j,k'} \right\rangle = \int_{-\infty}^{\infty} 2^j \varphi(2^j x - k)\overline{\varphi(2^j x - k')}\, dx$$
$$= \int_{-\infty}^{\infty} \varphi(y - k)\overline{\varphi(y - k')}\, dy$$
$$= \langle \varphi_{0,k}, \varphi_{0,k'} \rangle = \begin{cases} 0 & \text{if } k \neq k' \\ 1 & \text{if } k = k', \text{ by Equation 5.1} \end{cases}.$$

Now suppose $f(x) \in V_j$, then $f(2^{-j}x) \in V_0$, and so

$$f(2^{-j}x) = \sum_{k \in \mathbb{Z}} c_k \varphi(x - k)$$

in $L^2(\mathbb{R})$, and for some sequence $\{c_k\}_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$. But rescaling this, we just get that

$$f(x) = \sum_{k \in \mathbb{Z}} c_k 2^{-j/2} \varphi_{j,k}(x),$$

again, in $L^2(\mathbb{R})$. Finally, axiom (v) just tells us that these resolution spaces are closed under integer translations of functions.

The next example will come up quite often as we go along this section.

Example 34. Let V_j be the class of functions $f \in L^2(\mathbb{R})$ that are constant on every interval of the form $[2^{-j}n, 2^{-j}(n+1)), n \in \mathbb{Z}$. We will now show that $\{V_j\}_{j\in\mathbb{Z}}$ is a Multiresolution analysis. For any $f \in V_j$, we have that

$$f(x) = \sum_{n \in \mathbb{Z}} a_n \chi_{[n,n+1)}(2^j x)$$

in $L^2(\mathbb{R})$. Computing the norm we get

$$\|f\|_2 = \sum_{n\in\mathbb{Z}} \int_{2^{-j}n}^{2^{-j}(n+1)} |a_k|^2 \, dx = 2^{-j} \sum_{n\in\mathbb{Z}} |a_k|^2 < \infty.$$

This tells us there is a bijection T from V_j into $\ell^2(\mathbb{Z})$, such that to each $f(x) = \sum_{n\in\mathbb{Z}} a_n \chi_{[n,n+1)}(2^j x)$, T acts on f via

$$T(f) = \{a_n\}_{n\in\mathbb{Z}},$$

and further,

$$\|T(f)\|_{\ell^2} = 2^{j/2} \|f\|_{L^2}. \tag{10}$$

Now suppose $\{f_k\}_{k\in\mathbb{N}}$ is a Cauchy sequence in V_j, then $\{T(f_k)\}_{k\in\mathbb{N}}$ is also Cauchy in ℓ^2 by the equality in 10, giving us a limit $\{a_n\}_{n\in\mathbb{Z}} \in \ell^2$, meaning that

$$T^{-1}(\{a_n\}_{n\in\mathbb{Z}}) = \sum_{n\in\mathbb{Z}} a_n \chi_{[2^{-j}n, 2^{-j}(n+1))}(x) = \lim_{k\to\infty} f_k(x) \in V_j,$$

by the continuity of T. Axiom (i),(vi) and (v) follow immediately from the definition of V_j, and since step functions are dense in $L^2(\mathbb{R})$, (ii) does as well (remember that any half-open interval is measurable). To prove (iii), we note that on for a function f to be in V_j for every $j \in \mathbb{Z}$, means in particular that f is constant on $[0, 2^{-j})$ and $[-2^{-j}, 0)$, for every j, but if $f \in \bigcap_{j \in \mathbb{Z}} V_j$, then it is also constant on all of $\mathbb{R}$, meaning it must be thee zero function, as non-zero constant functions aren't square-integrable. Finally, showing (vi), a natural choice of our scaling function is $\varphi = \chi_{[0,1)}$. Translating this function by integer values gives us a sequence of disjointly supported functions, meaning $\{\varphi_{0,k}\}_{k \in \mathbb{Z}}$ is orthonormal. For $f \in V_0$ then, f is again a simple function on each unit interval of the form $[n, n+1)$, meaning it can be represented at each point x as

$$f(x) = \sum_{k \in \mathbb{Z}} f(k)\varphi(x-k).$$

Computing the Fourier coefficients for f we get for every $k \in \mathbb{Z}$

$$c_k(f) = \left\langle f, \chi_{[0,1)} \right\rangle = \int_{-\infty}^{\infty} f(x)\chi_{[0,1)}(x-k)\,dx = \int_{k}^{k+1} f(x)\,dx = f(k).$$

In particular, we have that

$$\sum_{k \in \mathbb{Z}} |f(k)|^2 = \sum_{k \in \mathbb{Z}} \int_{k}^{k+1} |f(k)|^2 \; dx = \|f\|_2^2 < \infty,$$

and so by the converse of Parseval's identity (Equation (4)) , we have shown that $\{\varphi_{0,k}\}_{k \in \mathbb{Z}}$ is indeed an orthonormal basis of V_0.

There is also a weaker notion of a system being orthonormal in a Hilbert space, namely that of a *Riesz System*. Call a sequence $\{x_k\}_{k \in \mathbb{Z}} \in \ell^2(\mathbb{Z})$ in a vector space *finitely non-zero* if $x_n \neq 0$ only for a finite number of terms.

Definition 35. Suppose H is a Hilbert space. A sequence of vectors $\{\varphi_n\}_{n \in \mathbb{Z}}$ is called a *Riesz system* if there exists two constants $D \geq C > 0$, such that for all finitely non-zero sequences $\{a_n\}_{n \in \mathbb{Z}} \in \ell^2(\mathbb{Z})$, we have

$$C \sum_{n \in \mathbb{Z}} |a_n|^2 \leq \| \sum_{n \in \mathbb{Z}} a_n \varphi_n \|^2 \leq D \sum_{n \in \mathbb{Z}} |a_n|^2 .$$

Example 36. Let

$$\varphi(x) = (1 + |x|)\chi_{[-1,1]}(x) x \in \mathbb{R}.$$

We call this compactly supported function the *tent function*. We will show that $\{\varphi_{0,k}\}_{k \in \mathbb{Z}}$ makes up a Riesz system. We start off by noting that on each interval $[n, n+1]$, $\varphi_{0,n}$ and $\varphi_{0,n+1}$ are the only elements in the sequence with overlapping support. Therefore, let $\{a_k\}_{k \in \mathbb{Z}}$ be finitely non-zero. We have that

$$\begin{aligned}
\left\|\sum_{k\in\mathbb{Z}} a_k\varphi_{0,k}\right\|_2^2 &= \int_{-\infty}^{\infty}\left|\sum_{k\in\mathbb{Z}} a_k\varphi_{0,k}(x)\right|^2\,dx \\
&= \sum_{n\in\mathbb{Z}}\int_n^{n+1} |a_n\varphi_{0,n}(x)+a_{n+1}\varphi_{0,n+1}(x)|^2\,dx \\
&= \sum_{n\in\mathbb{Z}}\int_n^{n+1} |a_n(n+1-x)+a_{n+1}(x-n)|^2\,dx \\
&= \sum_{n\in\mathbb{Z}}\int_0^{1} |a_n(1-t)+a_{n+1}t|^2\,dt \\
&= \sum_{n\in\mathbb{Z}}\int_0^{1} |a_n|^2(1-t)^2+|a_{n+1}|^2t^2+2\mathrm{Re}(a_n\overline{a_{n+1}})(t-t^2)\,dt \\
&=\frac{1}{3}\sum_{n\in\mathbb{Z}}(|a_n|^2+|a_{n+1}|^2+\mathrm{Re}(a_n\overline{a_{n+1}})).
\end{aligned}$$

Since $2|\mathrm{Re}(a_n\overline{a_{n+1}})| \leq |\,|a_n|^2+|\overline{a_{n+1}}|^2\,| = |a_n|^2+|a_{n+1}|^2$, we get that

$$\left\|\sum_{k\in\mathbb{Z}} a_k\varphi_{0,k}\right\|_2^2 \leq \frac{1}{3}\sum_{k\in\mathbb{Z}}\frac{3}{2}\left(|a_n|^2+|a_{n+1}|^2\right) \leq \sum_{k\in\mathbb{Z}}|a_n|^2,$$

as well as

$$\left\|\sum_{k\in\mathbb{Z}} a_k\varphi_{0,k}\right\|_2^2 \geq \frac{1}{6}\sum_{k\in\mathbb{Z}}|a_k|^2+|a_{n+1}|^2 = \frac{1}{3}\sum_{k\in\mathbb{Z}}|a_k|^2.$$

Now it is not too hard to see that with $C=\frac{1}{3}$ and $D=1$, $\{\varphi_{0,k}\}_{k\in\mathbb{Z}}$ fulfills the Riesz condition.

Corollary 33.1. Any orthonormal sequence is a Riesz system with $C=D=1$. This follows directly from Parseval's identity.

With the help of the Fourier transform, we can put a sufficient condition on when a function's integer translates creates a Riesz system.

Theorem 34. Suppose $\varphi\in L^2(\mathbb{R})$. Then the sequence $\{\varphi_{0,k}\}_{k\in\mathbb{Z}}$ is a Riesz system if and only if

$$C\leq\sum_{k\in\mathbb{Z}}|\widehat{\varphi}(\xi+2k\pi)|^2\leq D \text{ for a.e. } \xi\in\mathbb{R}. \tag{11}$$

The series above is called a *periodization*, and it is not hard to show that it is 2π–periodic, and that it lives in $L^2_p(0,2\pi)$.

Proof. Let $\{a_k\}_{k\in\mathbb{Z}}$ be finitely non-zero, and define

$$\Phi(x)=\sum_{k\in\mathbb{Z}}a_k\varphi(x-k).$$

Then from identities we have developed for the Fourier transform, we have that

$$\widehat{\Phi}(\xi) = \sum_{k\in\mathbb{Z}} a_k e^{-ik\xi}\widehat{\varphi}(\xi).$$

Note that since a_k is finitely non-zero, we have that $m(\xi) = \sum_{k\in\mathbb{Z}} a_k e^{-ik\xi}$ is a trigonometric polynomial. In particular, $m(\xi)$ is 2π–periodic. From Plancherel's formula in Theorem 33.3, we have that

$$\begin{aligned} 2\pi\|\Phi\|_2^2 = \|\widehat{\Phi}(\xi)\|^2 &= \int_{\mathbb{R}} |m(\xi)|^2\,|\widehat{\varphi}(\xi)|^2\,d\xi = \sum_{n\in\mathbb{Z}} \int_{2n\pi}^{2(n+1)\pi} |m(\xi)|^2\,|\widehat{\varphi}(\xi)|^2\,d\xi \\ &= \sum_{n\in\mathbb{Z}} \int_0^{2\pi} |m(\xi)|^2\,|\widehat{\varphi}(\xi+2n\pi)|^2\,d\xi = \int_0^{2\pi} |m(\xi)|^2 \left(\sum_{n\in\mathbb{Z}} |\widehat{\varphi}(\xi+2n\pi)|^2\right) d\xi. \end{aligned}$$

Rewritten neatly, we get

$$\left\|\sum_{k\in\mathbb{Z}} a_k\varphi_{0,k}\right\|_2^2 = \frac{1}{2\pi}\int_0^{2\pi} |m(\xi)|^2 \left(\sum_{n\in\mathbb{Z}} |\widehat{\varphi}(\xi+2n\pi)|^2\right) d\xi.$$

Now assume that $\widehat{\varphi}$ satisfies (11). Since

$$\frac{1}{2\pi}\int_0^{2\pi} |m(\xi)|^2\,d\xi = \sum_{k\in\mathbb{Z}} |a_k|^2\,\frac{1}{2\pi}\int_0^{2\pi} |e^{-ik\xi}|^2\,d\xi = \sum_{k\in\mathbb{Z}} |a_k|^2 \text{ by Parseval's Theorem,}$$

we get that

$$C \le \sum_{n\in\mathbb{Z}} |\widehat{\varphi}(\xi+2n\pi)|^2 \le D \text{ almost everywhere}$$

becomes

$$C\sum_{k\in\mathbb{Z}} |a_k|^2 \le \|\Phi(x)\|_2^2 \le D\sum_{k\in\mathbb{Z}} |a_k|^2$$

which means that $\{\varphi_{0,k}\}_{k\in\mathbb{Z}}$ is a Riesz system. To show the other direction, suppose φ satisfies the Riesz condition, then the right-hand side of (5.1) satisfies

$$C\int_0^{2\pi} |m(\xi)|^2\,d\xi \le \int_0^{2\pi} |m(\xi)|^2 \left(\sum_{n\in\mathbb{Z}} |\widehat{\varphi}(\xi+2n\pi)|^2\right) d\xi \le D\int_0^{2\pi} |m(\xi)|^2\,d\xi$$

for every trigonometric polynomial m. Via a very slick trick that involves approximating the characteristic function over an interval $[\xi_0-\varepsilon,\xi_0+\varepsilon] \subset (0,2\pi)$ in a way that let's us pass a limit we get that

$$C \le \frac{1}{2\varepsilon}\int_{\xi_0-\varepsilon}^{\xi_0+\varepsilon} \sum_{n\in\mathbb{Z}} |\widehat{\varphi}(\xi+2n\pi)|^2\,d\xi \le D.$$

Then we can be slick once more, and pretend we can let $\varepsilon \to 0$, and apply our differentiation theorem we forgot to write down, and we get our desired inequality. I don't know how to justify all steps in this proof! ☺

A corollary of this, is that integer translations of $\varphi \in L^2(\mathbb{R})$ generating an orthonormal sequence in $L^2(\mathbb{R})$ is equivalent to saying that

$$\sum_{k\in\mathbb{Z}} |\widehat{\varphi}(\xi + 2k\pi)|^2 = 1 \text{ for a.e. } \xi \in \mathbb{R}. \tag{12}$$

Example 37. We know already that the Scaling function in Example 34 for the Haar system in Example 34 generates an orthonormal sequence in $L^2(\mathbb{R})$, meaning that

$$\sum_{n\in\mathbb{Z}} \left|\widehat{\chi_{[0,1)}}(\xi + 2n\pi)\right|^2 = 1 = \sum_{n\in\mathbb{Z}} e^{-i\xi/2} \operatorname{sinc}^2(\xi/2 + 2n\pi) \text{ for every } \xi \in \mathbb{R} \setminus 2\pi\mathbb{Z},$$

where the right-hand side was computed back in Example 7.

Example 38. On a completely unrelated note, take $\chi_{(-\pi,\pi)}$. Then, as we have seen just now, the Fourier transform is

$$\widehat{\chi}_{(-\pi,\pi)}(\xi) = 2\pi e^{-i\xi(\pi-\pi)/2} \operatorname{sinc}(\pi\xi) = 2\pi \operatorname{sinc}(\pi\xi),\ \xi \in \mathbb{R} \setminus 0.$$

Via the inversion formula, we have that

$$\overline{\mathscr{F}}\left(\widehat{\chi}_{(-\pi,\pi)}\right) = \chi_{(-\pi,\pi)}, \text{ for a.e. } \xi \in \mathbb{R}.$$

Setting $\varphi = \operatorname{sinc}(\pi\xi)$, observe that

$$\sum_{k\in\mathbb{Z}} |\widehat{\varphi}(\xi + 2k\pi)|^2 = \sum_{k\in\mathbb{Z}} \chi(\xi + 2k\pi) = 1 \text{ for a.e. } \xi \in \mathbb{R},$$

since each summand is disjoint support from the others. This tells us that the sequence $\{\varphi_{0,k}\}_{k\in\mathbb{Z}}$ is orthonormal in $L^2(\mathbb{R})$. Later we will see a Multiresolution analysis coupled with φ as its scaling function, and the sequence $\{\varphi_{0,k}\}_{k\in\mathbb{Z}}$ called the *Shannon system*.

It appears to be the case that we can omit our sequences to be finitely generated in our definition of a Riesz system, and nothing changes! We just need a_k to live in $\ell^2(\mathbb{Z})$. The proof of this can be found in [TUR15, p. 10]. The next lemma just tells us that the closed linear span of a Riesz system yields us the same functions that an orthonormal basis would.

Lemma 35. Let $\Phi = \{\varphi_{0,k}\}_{k\in\mathbb{Z}}$ be a Riesz system in $L^2(\mathbb{R})$. Then for any $f \in L^2(\mathbb{R})$, we have that f is in the closed linear span of Φ if and only if

$$f(x) = \sum_{k\in\mathbb{Z}} a_k \varphi_{0,k}(x) \tag{13}$$

in $L^2(\mathbb{R})$ for some sequence $\{a_k\}_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$.

Proof. To prove sufficiency, suppose f yielded the representation given in (13) in $L^2(\mathbb{R})$. Then taking the partial sums of the series representation yields us a sequence f_n in the span (Φ) converging to f in $L^2(\mathbb{R})$. Now consider when f u=is is in the closed linear span of $\{\varphi_{0,k}\}_{k\in\mathbb{Z}}$. Then there is a sequence $\{f_n\}_{n\in\mathbb{Z}}$ completely contained in the span of Φ converging to f in $L^2(\mathbb{R})$. In particular,

$$f_n(x) = \sum_{k\in\mathbb{Z}} a_k^{(n)} \varphi_{0,k}(x), x \in \mathbb{R},$$

and $\{a_k^{(n)}\}_{k\in\mathbb{Z}}$ is finitely non-zero for $n = 1, 2, 3, \ldots$. For each n, set $a_n = \{a_k^n\}_{k\in\mathbb{Z}}$. Then $a_n - a_m$ is the associated sequence to the function $f_m - f_n$, that is also in the linear span of Φ Therefore, there exists some $C > 0$ such that

$$C\|a - m - a_n\|_{\ell^2}^2 \leq \|f_m - f_n\|_{L^2}^2,$$

meaning that $\{a_n\}_{n\in\mathbb{Z}}$ is Cauchy, and so by the completeness of $\ell^2(\mathbb{Z})$, convergences to some sequence $a = \{a'_k\}_{k\in\mathbb{Z}}$ in $\ell^2(\mathbb{Z})$. Let

$$S_K(x) = \sum_{|k|\leq K} a'_k \varphi_{0,k}(x),\ x \in \mathbb{R},\ K = 0, 1, \ldots.$$

Then if $\|f - S_K\|_{L^2}$ can be made as small as possible, the representation in (13) makes sense in $L^2(\mathbb{R})$. Therefore, let $\varepsilon > 0$, choose n_0 large enough so that

$$\|f - f_n\|_{L^2} < \varepsilon \text{ and } \|a - a_n\|_{\ell^2} < \varepsilon \text{ whenever } n \geq n_0.$$

Further, choose K_0 large enough so that $\{a'_k\}_{|k|\leq K_0} = a$. By the triangle inequality, we have that

$$\begin{aligned}
\|f - S_K\|_{L^2} \leq& \|f - f_n\|_{L^2} + \|f_n - S_K\|_{L^2} \\
\leq& \varepsilon + \left\| \sum_{|k|\leq K} (a_k^{(n)} - a'_k)\varphi_{0,k} \right\|_{L^2} \\
\leq& \varepsilon + \sqrt{D}\|a_n - a\|_{\ell^2} \leq (1 + \sqrt{D})\varepsilon.
\end{aligned}$$

Since ε was chosen arbitrarily, we have that the representation of f in (13) is true in $L^2(\mathbb{R})$. ☺

Remark 6. If $\{\varphi_{0,k}\}_{k\in\mathbb{Z}}$ is a Riesz system in $L^2(\mathbb{R})$. Then, $\{\varphi_{j,k}\}_{k\in\mathbb{Z}}$ is also a Riesz system for any $j \in \mathbb{Z}$, since

$$\begin{aligned}
\left\| \sum_{k\in\mathbb{Z}} a_k \varphi_{j,k} \right\|_2^2 &= \int_{\mathbb{R}} \left| \sum_{k\in\mathbb{Z}} a_k 2^{j/2} \varphi(2^j x - k) \right|^2 dx \\
&= \int_{\mathbb{R}} \left| \sum_{k\in\mathbb{Z}} a_k \varphi(y - k) \right| dy, y = 2^j x \\
&= \left\| \sum_{k\in\mathbb{Z}} a_k \varphi_{0,k} \right\|_2^2
\end{aligned}$$

for any finitely non-zero sequence $\{a_k\}_{k\in\mathbb{Z}}$. Therefore, any closed linear span of a Riesz system $\{\varphi_{j,k}\}_{k\in\mathbb{Z}}$ generates the subspace of all functions f in $L^2(\mathbb{R})$ such that

$$f(x) = \sum_{k\in\mathbb{Z}} a_k \varphi_{j,k}(x)$$

in $L^2(\mathbb{R})$, for some sequence $\{a_k\}_{k\in\mathbb{Z}}$ in $\ell^2(\mathbb{Z})$.

The next theorem tells us, given a Riesz system, we can always construct an orthonormal basis with the same (closed) span as the original Riesz system.

Theorem 36. Suppose $\{\varphi_{0,k}\}_{k\in\mathbb{Z}}$ is a Riesz system in $L^2(\mathbb{R})$. Then there exists a sequence $\{a_k\}_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$ and a function $\Phi \in L^2(\mathbb{R})$ with

$$\Phi(x) = \sum_{k\in\mathbb{Z}} a_k \varphi_{0,k}(x),$$

such that $\{\Phi_{0,k}\}_{k\in\mathbb{Z}}$ is an orthonormal system in $L^2(\mathbb{R})$, and

$$\overline{\text{span}(\{\Phi_{0,k}\})} = \overline{\text{span}(\{\varphi_{0,k}\})}.$$

Proof. To get our sequence $\{a_k\}_{k\in\mathbb{Z}}$, we exploit that the Fourier transform does seem to give us a sequence when considering a Riesz system. Therefore, put

$$\widehat{\Phi}(\xi) = \frac{\widehat{\varphi}(\xi)}{\left(\sum_{n\in\mathbb{Z}} |\widehat{\varphi}(\xi + 2n\pi)|^2\right)^{1/2}},\ \xi \in \mathbb{R}.$$

Denote the denominator in the equation above by $m(\xi)$, then $\widehat{\Phi}(\xi) = (m^{-1} \cdot \widehat{\varphi})(\xi)$. By Theorem 34, we have that $m(\xi)$ is bounded below by some $C > 0$, and also bounded above by some constant D for almost every $\xi \in \mathbb{R}$, and so $\widehat{\Phi}$ is an element of $L^2(\mathbb{R})$. Therefore, $\Phi(x) = \overline{\mathscr{F}}(\widehat{\Phi})$ a.e. on $\mathbb{R}$. Now to prove $\{\Phi_{0,k}\}_{k\in\mathbb{Z}}$ is an orthonormal sequence, we just need to check that

$$\sum_{k\in\mathbb{Z}} |\Phi(\xi + 2k\pi)|^2 = 1 \text{ for a.e. } \xi \in \mathbb{R}.$$

And this is true! Since $m^{-1}(\xi)$ is 2π–periodic. Further, $m^{-1}(\xi)$ is in $L^2_p(0, 2\pi)$ (remember that $m(\xi)$ is bounded strictly above and below), there exists some sequence $\{a_k\}_{k\in\mathbb{Z}}$ in $\ell^2(\mathbb{Z})$ such that

$$m^{-1}(\xi) = \sum_{k\in\mathbb{Z}} a_k e^{-ik\xi}.$$

The inverse Fourier transform on $\widehat{\Phi}(\xi)$ is then

$$\Phi(x) = \int_{\mathbb{R}} m^{-1}(\xi)\widehat{\varphi}(\xi)e^{i\xi x}\, dx = \sum_{k\in\mathbb{Z}} a_k \int_{\mathbb{R}} \widehat{\varphi}(\xi)e^{i\xi(x-k)}\, dx = \sum_{k\in\mathbb{Z}} a_k\varphi_{0,k}(x).$$

To show the closed linear span of $\{\Phi_{0,k}\}_{k\in\mathbb{Z}}$ and $\{\varphi_{0,k}\}_{k\in\mathbb{Z}}$ coincide, we want to consider whether or not

$$\sum_{k\in\mathbb{Z}} b_k\varphi_{0,k} = \sum_{k\in\mathbb{Z}} c_k\Phi_{0,k} \tag{14}$$

always has a solution if $\{b_k\}_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$ is unknown or $\{c_k\}_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$ is unknown. If we take Fourier transforms on both sides, our equation becomes

$$B(\xi)\widehat{\varphi}(\xi) = C(\xi)\widehat{\Phi}(\xi) = C(\xi)m^{-1}(\xi)\widehat{\varphi(\xi)},\ \xi \in \mathbb{R},$$

where

$$B(\xi) = \sum_{k\in\mathbb{Z}} b_k e^{-ik\xi} \quad \text{and} \quad C(\xi) = \sum_{k\in\mathbb{Z}} c_k e^{-ik\xi}.$$

Note that $B(\xi)/m(\xi)$ and $C(\xi)m(\xi)$ are both in $L^2_p(0, 2\pi)$ by a similar argument as before. Therefore, assume $\{b_k\}_{k\in\mathbb{Z}}$ is known, then we can solve Equation (14) by taking $C(\xi)$ to be the Fourier series of $\frac{B(\xi)}{m(\xi)}$, meaning we choose

$$c_k = \left\langle \frac{B}{m}, e^{ik\xi} \right\rangle,$$

which shows that the closed linear span of $\{\Phi_{0,k}\}_{k\in\mathbb{Z}}$ is contained in that of $\{\varphi_{0,k}\}_{k\in\mathbb{Z}}$. To get the reverse inclusion, suppose each b_k is unknown, then we repeat the same argument but this time for $B(\xi) = m(\xi) \cdot C(\xi)$. ☺

5.2 Scaling Equation and the structure constants

As we saw earlier, $\varphi \in L^2(\mathbb{R})$ being the scaling fucntion for a multiresolution analysis $\{V_j\}_{j\in\mathbb{Z}}$, means that each set $\{\varphi_{j,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis for V_j. In particular, any function f in V_1 can be written as

$$f(x) = \sum_{k\in\mathbb{Z}} a_k \varphi_{1,k}(x)$$

where

$$a_k = \langle f, \varphi_{1,k} \rangle = 2^{1/2} \int_{\mathbb{R}} f(x)\overline{\varphi(2x-k)}\, dx. \tag{15}$$

Since $V_j \subset V_{j+1}$, even the scaling function itself is of the form

$$\varphi(x) = \sum_{k\in\mathbb{Z}} c_k \varphi_{1,k}(x) \tag{16}$$

in $L^2(\mathbb{R})$. We will see that this representation plays an important role for when we want to generate a Wavelet basis out of our Multiresolution analysis, and so we will call Equation (16) the *scaling equation*, and the numbers c_k the *structure constants*.

Example 39. If we go back to our example of the Haar system, with its scaling function being $\varphi = \chi_{[0,1)}$, we note that

$$\varphi(x) = \varphi(2x) + \varphi(2x-1), x \in \mathbb{R},$$

and so the scaling equation for this system becomes

$$\varphi(x) = \frac{1}{\sqrt{2}}\varphi_{0,1}(x) + \frac{1}{\sqrt{2}}\varphi_{1,1}(x),$$

with structure constants $c_0 = c_1 = \frac{1}{\sqrt{2}}$, and $c_k = 0$ for $k \neq 0, 1$.

The next theorem gives us an orthogonality condition that the structure constants satisfy.

Theorem 37. Let $\{c_k\}_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$ be a sequence of structure constants for a multiresolution analysis. Then

$$\sum_{k\in\mathbb{Z}} c_k \overline{c_{2m+k}} = \begin{cases} 1 \text{ if } & m = 0 \\ 0 \text{ if } & m \neq 0 \end{cases}.$$

Proof. We know that $\varphi(x) = \sum_{k\in\mathbb{Z}} c_k \varphi_{1,k}(x)$, and that $\varphi_{0,-m}(x) = \sum_{k\in\mathbb{Z}} c_k^{(m)} \varphi_{1,k}(x)$, where

$$\begin{aligned} c_k^{(m)} = \langle \varphi_{0,-m}, \varphi_{1,k} \rangle &= \int_{\mathbb{R}} \varphi_{0,-m}(x) \cdot \overline{\varphi_{1,k}}(x)\, dx \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{im\xi} \widehat{\varphi}(\xi) \cdot \overline{\widehat{\varphi_{1,k}}}(\xi)\, d\xi \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{im\xi} \widehat{\varphi} \cdot 2^{-1/2} e^{ik\xi/2} \overline{\widehat{\varphi}}\,(\xi/2)\; d\xi \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{\varphi}(\xi) \cdot 2^{-1/2} e^{i\xi(m+k/2)} \overline{\widehat{\varphi}}(\xi/2)\, d\xi \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{\varphi}(\xi) \cdot \overline{\widehat{\varphi_{1,2m+k}}}(\xi)\, d\xi \\ &= \int_{\mathbb{R}} \varphi(x) \cdot \overline{\varphi_{1,2m+k}}(x)\, dx \\ &= \langle \varphi, \varphi_{1,2m+k} \rangle = c_{2m+k}. \end{aligned}$$

The second step is just an application of Plancherel (Theorem 33.3), and in the third step, we make use of our identity in (9). Now we also know that,

$$\left.\begin{matrix} 1 & \text{if } m = 0, \\ 0 & \text{if } m \neq 0 \end{matrix}\right\} = \langle \varphi, \varphi_{0,-m} \rangle = \sum_{k\in\mathbb{Z}} c_k \overline{c_k^{(m)}} \|\varphi_{1,k}\|_2^2 = \sum_{k\in\mathbb{Z}} c_k \overline{c_{2m+k}},$$

concluding the proof. ☺

The next theorem highlights the *filter* of a function, and an equality called the *filter identity*.

Theorem 38. Suppose that $\varphi \in L^2(\mathbb{R})$ is the scaling function for a multiresolution analysis $\{V_j\}_{j\in\mathbb{Z}}$. Then, for every $f \in V_1$, there exists a function $m_f \in L_p^2(0, 2\pi)$ such that

$$\widehat{f}(\xi) = m_f\left(\frac{\xi}{2}\right) \widehat{\varphi}\left(\frac{\xi}{2}\right) \quad \text{for a.e. } \xi \in \mathbb{R}. \tag{17}$$

The function m_f is given by

$$m_f(\xi) = \frac{1}{\sqrt{2}} \sum_{k\in\mathbb{Z}} a_k e^{-ik\xi}, \xi \in \mathbb{R}. \tag{18}$$

where a_k is the Fourier coefficient of f in the basis $\{\varphi_{1,k}\}_{k\in\mathbb{Z}}$.

Proof. Since $\{\varphi_{1,k}\}_{k\in\mathbb{Z}}$ is a basis for $f \in V_1$, we have

$$f(x) = \sum_{k\in\mathbb{Z}} a_k \varphi_{1,k}(x) \text{ in } L^2(\mathbb{R}).$$

Applying the Fourier transform and Equation (9), we get

$$\widehat{f}(\xi) = \sum_{k\in\mathbb{Z}} a_k \cdot \frac{1}{\sqrt{2}} e^{-ik\xi/2} \widehat{\varphi}\left(\frac{\xi}{2}\right) = m_f\left(\frac{\xi}{2}\right) \widehat{\varphi}\left(\frac{\xi}{2}\right).$$

Note that $m_f(\xi)$ is just a Fourier series in $L_p^2\,(0, 2\pi)$ since $\{a_k\}_{k\in\mathbb{Z}}$ lives in $\ell^2(\mathbb{Z})$ by construction. ☺

Remark 7. Note that any function $\varphi \in L^2(\mathbb{R})$ for which $\{\varphi_{0,k}\}_{k\in\mathbb{Z}}$ is an orthonormal sequence (and so in particular, $\{\varphi_{1,k}\}_{k\in\mathbb{Z}}$ is orthonormal), it holds that if φ satisfies a filter identity given in (17), it also satisfies a scaling equation. That is, the coefficients in

$$m_\varphi(\xi) = \frac{1}{\sqrt{2}} \sum_{k\in\mathbb{Z}} a_k e^{-ik\xi},$$

are in $\ell^2(\mathbb{Z})$ since $m_\varphi \in L^2(0, 2\pi)$, and so the sequence

$$\sum_{|k|<n} h_k \varphi_{1,k} \text{ converges in } L^2\text{-norm},$$

by Lemma 21. As we have seen earlier in the proof of Theorem 38, the Fourier transform of the above is precisely the filter identity, and so

$$\varphi = \sum_{k\in\mathbb{Z}} h_k \varphi_{1,k} \text{ in } L^2(\mathbb{R}),$$

meaning that the coefficients h_k are precisely the structure constants of φ.

Example 40. Returning to our Haar system, with the scaling equation being

$$\varphi(x) = \frac{1}{\sqrt{2}}\varphi_{0,1}(x) + \frac{1}{\sqrt{2}}\varphi_{1,1}(x),\ x \in \mathbb{R},$$

we have that

$$\widehat{\varphi_{0,1}}(\xi) = 2^{-1/2}\widehat{\varphi}(\xi/2) \quad \text{and} \quad \widehat{\varphi_{1,1}}(\xi) = 2^{-1/2}e^{-i\xi/2}\widehat{\varphi}(\xi/2),$$

so the filter identity is

$$\widehat{\varphi}(\xi) = m_\varphi(\xi/2)\widehat{\varphi}(\xi/2), \text{ where } m_\varphi(\xi) = \frac{1}{2}(1 + e^{-i\xi}),$$

for each $\xi \in \mathbb{R}$.

Example 41. For the Shannon system that we saw in Example 38, generated by the function $\operatorname{sinc}(\pi x)$, we have that

$$\widehat{\varphi}(\xi) = \chi_{(-\pi,\pi)}(\xi),\ \xi \in \mathbb{R}.$$

The filter identity becomes

$$\chi_{(-\pi,\pi)}(\xi) = m_\varphi(\xi/2)\chi_{(-\pi,\pi)}(\xi/2) = m_\varphi(\xi/2)\chi_{(-2\pi,2\pi)}(\xi), \text{ for } \xi \in \mathbb{R}.$$

m_φ therefore is $\chi_{(-\pi/2,\pi/2)}$, and thankfully, we already know its Fourier coefficients, namely

$$\widehat{m_\varphi}(k) = \frac{1}{2\pi}\int_{-\pi/2}^{\pi/2} e^{ik\xi}\, d\xi = \begin{cases} \frac{\operatorname{sinc}(k\pi/2)}{2} & \text{for } k \neq 0 \\ \frac{1}{2} & \text{for } k = 0 \end{cases},$$

and so the structure constants are

$$c_k = \frac{\operatorname{sinc}(k\pi/2)}{\sqrt{2}},\ k \neq 0 \quad \text{and} \quad c_0 = \frac{1}{\sqrt{2}}.$$

5.3 Generating a Multiresolution Analysis

We have seen some nice examples of Multiresolution analyses, but often, this is a cumbersome task. The next theorem lets us construct one more easily, given we have a proper candidate for a scaling function.

Theorem 39. Let $\varphi \in L^2(\mathbb{R})$ satisfy the following

1. $\{\varphi_{0,k}\}_{k\in\mathbb{Z}}$ is an orthonormal sequence in $L^2(\mathbb{R})$;
2. φ satisfies a scaling equation, namely
$$\varphi(x) = \sum_{k\in\mathbb{Z}} c_k \varphi_{1,k}(x)$$
in $L^2(\mathbb{R})$ for some sequence $\{c_k\}_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$;
3. $\widehat{\varphi}$ is continuous at 0 with $|\widehat{\varphi}(0)| = 1$.

Then the sequence $\{V_j\}_{j\in\mathbb{Z}}$ – with each V_j defined as the closed linear span of $\{\varphi_{j,k}\}_{k\in\mathbb{Z}}$ – is a multiresolution analysis of $L^2(\mathbb{R})$.

Proof. We start off by noting that for each $j \in \mathbb{Z}$, $\{\varphi_{j,k}\}_{k\in\mathbb{Z}}$ are also orthonormal (Remark 6), and so each V_j only consists of functions of the form
$$f = \sum_{k\in\mathbb{Z}} a_k \varphi_{j,k}(x) \quad \text{in } L^2(\mathbb{R}), \text{ where } \{a_k\}_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z}).$$

Further, let $P_j f$ denote the orthogonal projection of a function $f \in L^2(\mathbb{R})$ on V_j for $j \in \mathbb{Z}$. That is,
$$P_j f = \sum_{k\in\mathbb{Z}} \left\langle f, \varphi_{j,k} \right\rangle \varphi_{j,k}.$$

We need to check that all the properties of Definition 33 are satisfied.

(i) Suppose that $f \in V_j$, that is
$$f(x) = \sum_{k\in\mathbb{Z}} a_k \varphi_{j,k}(x) = \sum_{k\in\mathbb{Z}} (2^{j/2} a_k \varphi(2^j x - k)$$
in $L^2(\mathbb{R})$, and for some $\{a_k\}_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$. By the scaling equation, we have that
$$\varphi(2^j x - k) = \sum_{l\in\mathbb{Z}} c_l \varphi_{1,l}(2^j x - k) = \sum_{l\in\mathbb{Z}} 2^{1/2} c_l \varphi(2^{j+1} x - 2k - l).$$
Inserting this into our previous identity, we get that
$$f(x) = \sum_{k,l\in\mathbb{Z}} (2^{(j+1)/2} a_k c_l) \varphi(2^{j+1} x - 2k - l).$$
For each $n \in \mathbb{Z}$, we can set $d_n = \sum_{2k+l=n} a_k c_l$, and get the representation
$$f(x) = \sum_{n\in\mathbb{Z}} d_n \varphi_{j+1,n}.$$
This shows that $f \in V_{j+1}$.

(ii) Here we will want to show that for $f \in L^2(\mathbb{R})$, $P_j f \to f$ in $L^2(\mathbb{R})$ as $j \to \infty$. This will imply that the $\bigcup_{j\in\mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R})$. Let $f \in L^2(\mathbb{R})$, and $\{f_n\}_{n\geq 1}$ be a sequence in $\mathscr{S}$ converging to f in $L^2(\mathbb{R})$. Then we have that

$$\begin{aligned}\|f - P_j f\|_2 \leq& \|f - f_n\|_2 + \|f_n - P_j f_n\|_2 + \|P_j(f_n - f)\|_2 \\ \leq& 2\|f - f_n\|_2 + \|f_n - P_j f_n\|_2,\end{aligned}$$

since the norm of a projection of a function can only be less or equal to the norm of said function (recall that projections are linear, and thus continuous maps). What we can see from the above inequality, is that if we make n large enough, we can get the first term as small as we want. This means that as j increases, we only need to check that $P_j f \to f$ in $L^2(\mathbb{R})$ for $f \in \mathscr{S}$. Or in other words, (because we will use this later), we want to show the statement for $\widehat{g} \in C_c^\infty$, with $\text{supp}(g) \subset [-R, R]$ for $R > 0$. By Theorem 30.1, we get $g \in \mathscr{S}$). From Pythagoras' identity, we know that

$$\|g - P_j g\|_2^2 = \|g\|_2^2 - \|P_j g\|_2^2.$$

therefore we only need to show that $\|P_j g\|_2 \to \|g\|_2$ as $j \to \infty$. Suppose j is large enough as to satisfy $2^{-j}R \leq \pi$. Since $\{\varphi_{j,k}\}_{k\in\mathbb{Z}}$ is an orthonormal system in V_j, Parseval's Theorem applies, and so we have

$$\|P_j g\|_2^2 = \sum_{k\in\mathbb{Z}} \left|\left\langle g, \varphi_{j,k}\right\rangle\right|^2.$$

To each term, Plancherel's Theorem applies and so we get

$$\sum_{k\in\mathbb{Z}} \left|\left\langle g, \varphi_{j,k}\right\rangle\right|^2 = \frac{1}{(2\pi)^2} \sum_{k\in\mathbb{Z}} \left|\int_{-R}^{R} \widehat{g}(\xi)\overline{\widehat{\varphi_{j,k}}(\xi)}\, d\xi\right|^2.$$

Now according to identity (9), and a change of variables $2^{-j}\xi = \omega$

$$\begin{aligned}\int_{-R}^{R} \widehat{g}(\xi)\overline{\widehat{\varphi_{j,k}}(\xi)}\, d\xi =& 2^{-j/2} \int_{-R}^{R} \widehat{g}(\xi) e^{i2^{-j}k\xi}\overline{\widehat{\varphi}(2^{-j}\xi)}\, d\xi \\ =& 2^{j/2} \int_{-2^{-j}R}^{2^{-j}R} \widehat{g}(2^j\omega)\overline{\widehat{\varphi}(\omega)} e^{ik\omega}\, d\omega \\ =& 2^{j/2} \int_{-\pi}^{\pi} \widehat{g}(2^j\omega)\overline{\widehat{\varphi}(\omega)} e^{ik\omega}\, d\omega.\end{aligned}$$

Here in the last step we can see that this is just a projection of $h(\omega) = \widehat{g}(2^j\omega) \cdot \overline{\widehat{\varphi}(\omega)}$ onto the basis vector $e^{-ik\omega}$ in $L_p^2(-\pi, \pi)$. From this we notice that h belongs to $L_p^2\,(-\pi, \pi)$. Therefore we can use Parseval once again and yield that

$$\begin{aligned}\|P_j g\|_2^2 =& \frac{2^j}{(2\pi)^2} \sum_{k\in\mathbb{Z}} \left|\int_{-\pi}^{\pi} h(\omega) e^{ik\omega}\, d\omega\right|^2 \\ =& \frac{2^j}{2\pi} \int_{-\pi}^{\pi} \left|\widehat{g}(2^j\omega)\right|^2 |\widehat{\varphi}(\omega)|^2\, d\omega \\ =& \frac{1}{2\pi} \int_{-R}^{R} |\widehat{g}(\xi)|^2 \left|\widehat{\varphi}(2^{-j}\xi)\right|^2\, d\xi.\end{aligned}$$

Since we assume that $\widehat{\varphi}$ is continuous at 0, we note that $\widehat{\varphi}(2^{-j}\xi)$ converges uniformly to the constant function 1 on $[-R, R]$ as $j \to \infty$ (just choose j such that $|2^{-j}R|$ is small enough. Then any $|x| \leq R$) This finally gives us that

$$\|P_j g\|_2^2 \to \frac{1}{2\pi} \int_{-R}^{R} |\widehat{g}(\xi)|^2 \, d\xi = \int_{\mathbb{R}} |g(x)|^2 \, dx$$

(iii) If we show that $P_j f \to 0$ in $L^2(\mathbb{R})$ as $j \to -\infty$, we will also have shown that $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$, since if $f \in \bigcap_{j\in\mathbb{Z}} V_j$, then for every $j \in \mathbb{Z}$, $P_j f = f$ and so taking limits will give us that f must be the zero function. Let $f_r = f \cdot \chi_{[-r,r]}$. We start off by bounding $\|P_j f\|_2$ by $\|P_j f_r\|_2$, then if we can show that $\|P_j f_r\|_2 \to 0$ as $j \to -\infty$, we are done. By linearity, we have that

$$\|P_j f\|_2 \leq \|P_j(f - f_r)\|_2 + \|P_j f_r\|_2 \leq \|f - f_r\|_2 + \|P_j f_r\|_2.$$

The first term can be made as small as possible by letting r get large enough. And so we get that for some $R_0 > 0$,

$$\|P_j f\|_2 \leq \varepsilon + \|P_j f_r\|_2, \text{ for } r > R_0.$$

Therefore, if we show what we want for functions with compact support (in particular, supp$f \subset [-R, R]$ for some R large enough), then we are done. We know from Parseval's identity that

$$\|P_j f\|_2 = \sum_{k\in\mathbb{Z}} \left|\left\langle f, \varphi_{j,k} \right\rangle\right|^2 \leq \sum_{k\in\mathbb{Z}} \left(\int_{-R}^{R} |f| \, |\varphi_{j,k}| \, dx \right)^2.$$

By the Cauchy-Schwartz inequality, we have that

$$\|P_j f\|_2^2 \leq \sum_{k\in\mathbb{Z}} \int_{-R}^{R} |f(x)|^2 \, dx \int_{-R}^{R} 2^j \left|\varphi(2^j x - k)\right|^2 \, dx$$

A change of variables for the second integral gives us the right-hand side to equal

$$\sum_{k\in\mathbb{Z}} \|f\|_2^2 \int_{-2^j R - k}^{2^j R - k} |\varphi(y)|^2 \, dy.$$

Here we can make use of the Dominated Convergence theorem (Theorem 10), to get that this integral goes to zero, and so we end up with the fact that

$$\lim_{j\to-\infty} \|P_j f\|_2^2 = 0,$$

and we are done! We have shown that for any $f \in L^2(\mathbb{R})$, the projection of f onto V_j becomes the zero function as $j \to -\infty$. This itself implies that

$$\bigcap_{k\in\mathbb{Z}} V_j = \{0\}.$$

(iv) Suppose we have that $f \in V_j$, then

$$f(2^{-j}x) = \sum_{k\in\mathbb{Z}} a_k \varphi_{j,k}(2^{-j}x) = \sum_{k\in\mathbb{Z}} (2^{j/2} a_k) \varphi_{0,k}(x) \in V_0.$$

Conversely, if $f(2^{-j}x) \in V_0$, then

$$f(x) = \sum_{k \in \mathbb{Z}} a_k \varphi(2^j x - k) = \sum_{k \in \mathbb{Z}} (2^{j/2} a_k) \varphi_{j,k}(x) \in V_j,$$

and so we are done.

(v) is fairly straightforward, it just follows the same argument as in (iv). Note also, that in showing (iv), we have also shown that $\{\varphi_{j,k}\}$ is an orthonormal basis of V_j, so that (vi) follows. ☺

Example 42. We return to the Shannon system from Example 38, where $\varphi = \operatorname{sinc}(\pi\xi)$. We have already shown that $\{\varphi_{0,k}\}_{k \in \mathbb{Z}}$ is orthonormal in $L^2(\mathbb{R})$. We also know by Remark 7 that φ satisfies a scaling equation, and as a matter of fact, we have also computed the structure constants in Example 41. Lastly, the characteristic function on $[-\pi, \pi]$ is obviously continuous at 0, giving us that we have a multiresolution analysis of $L^2(\mathbb{R})$ generated by φ!

5.4 Constructing a Wavelet out of a multiresolution analysis

Definition 43. Given a multiresolution analysis $\{V_j\}_{j \in \mathbb{Z}}$ of $L^2(\mathbb{R})$, define W_j as the subspace solving the direct sum

$$V_{j+1} = V_j \oplus W_j.$$

That is, W_j is the orthogonal complement of V_j in V_{j+1}, for $j \in \mathbb{Z}$. We will call W_j a *detail space*.

Theorem 40. Suppose $\{V_j\}_{j \in \mathbb{Z}}$ is a multiresolution analysis of $L^2(\mathbb{R})$. Then for every $j \in \mathbb{Z}$, $f(x) \in W_j$ if and only if $f(2^{-j}x) \in W_0$.

Proof. Suppose $f(x)$ is nonzero in W_j. then f cannot be an element of V_j, and so $f(2^{-j}x) \notin V_0$. However, $f(2^{-j-1}x) \in V_0$, and so $f(2^{-j}x) \in V_1$, from this it follows that $f(2^{-j}x) \in W_0$. All the steps also hold in reverse by the scaling property of V_j. ☺

What Theorem 40 is saying is that $\{W_j\}_{j \in \mathbb{Z}}$ inherits the same scaling properties as $\{V_j\}_{j \in \mathbb{Z}}$.

Theorem 41. Suppose that $\{V_j\}_{j \in \mathbb{Z}}$ is a multiresolution analysis of $L^2(\mathbb{R})$. Then

$$L^2(\mathbb{R}) = \bigoplus_{j \in \mathbb{Z}} W_j.$$

Proof. We show that

$$L^2(\mathbb{R}) = V_0 \oplus \bigoplus_{j=0}^{\infty} W_j, \text{ and that } V_0 = \bigoplus_{j=1}^{\infty} W_{-j}.$$

Therefore, suppose $f \in L^2(\mathbb{R})$ and in the orthogonal complement of the right hand side. Then that means that $f \perp V_0$ and $f \perp W_j = V_{j+1} \ominus V_j$ for all $j \geq 0$. Since $\bigcup_{j \geq 1} V_{-j} \subset V_0$, from this it follows that $f \perp \overline{\bigcup_{j \in \mathbb{Z}}} = L^2(\mathbb{R})$. (It is not hard to show that an element orthogonal to a dense set is also orthogonal to the closure of it. It mostly follows from the continuity of inner products.) Now we want to show that

$$V_0 = \bigoplus_{j=1}^{\infty} W_{-j}.$$

Suppose $f \in V_0$ and is orthogonal to the right-hand side of the above. Then f is orthogonal to every W_{-j} for every $j \geq 1$. But then f must also be an element of every V_{-j+1}, and so in particular $f \in \bigcap_{j\geq 0} V_{-j} = \{0\}$ and we are done. ☺

Now we come to the point where given a multiresolution analysis $\{V_j\}_{j\in\mathbb{Z}}$ of $L^2(\mathbb{R})$ with φ as its scaling function, and another function $\psi \in L^2(\mathbb{R})$ for which $\{\psi_{0,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis of the detail space W_0, then it is not too hard to see thanks to the previous two results that $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ is a wavelet basis for $L^2(\mathbb{R})$. The fact that the sequence spans $L^2(\mathbb{R})$ just follows from Theorem 41, and the orthogonality follows from the following. Let j, l, k, k' be integers such that $j \neq l$, then $\psi_{j,k} \perp \psi_{l,k'}$. Since inner products are conjugate symmetric, we can safely assume $j \geq l$, and note that $\psi_{j,k} \in W_j \perp V_j \supseteq V_l \ni \psi_{l,k'}$. Therefore, $\psi_{j,k}$ and $\psi_{l,k'}$ must be orthogonal.

Now is the final stretch. We almost have everything we need to construct a Wavelet basis $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ of $L^2(\mathbb{R})$, out of a multiresolution analysis. To aid with that, the following will be referred to in the next theorem. Let $\varphi \in V_0$ be the scaling function for the multiresolution analysis of $L^2(\mathbb{R})$, $\{V_j\}_{j\in\mathbb{Z}}$. For some some $\psi \in W_0$, we have by Equation (17), that the function $m_\psi(\xi) \in L^2_p(0, 2\pi)$ is the associated filter of ψ satisfying

$$\widehat{\psi}(\xi) = m_\psi(\xi/2)\widehat{\varphi}(\xi/2) \text{ for a.e. } \xi \in \mathbb{R}. \tag{19}$$

the matrix M defined via

$$M(\xi) = \begin{pmatrix} m_\varphi(\xi) & m_\psi(\xi) \\ m_\varphi(\xi+\pi) & m_\psi(\xi+\pi) \end{pmatrix}, \quad \xi \in \mathbb{R}. \tag{20}$$

It is worth reminding the reader that an $n \times n$ matrix U with complex entries $(U)_{j,k} \in \mathbb{C}$ is *Unitary*, if

$$U^*U = UU^* = I,$$

where I is the identity matrix, and U^* is the conjugate transpose of U. That is, $(U^*)_{j,k} = \overline{(U)_{k,j}}$.

Theorem 42. Suppose that $\{V_j\}_{j\in\mathbb{Z}}$ is a multiresolution analysis of $L^2(\mathbb{R})$ with scaling function φ.

(a) If $\{\psi_{0,k}\}_{k\in\mathbb{Z}}$ is an orthonormal sequence in W_0, then $M(\xi)$ as defined in (20) is unitary almost everywhere.

(b) On the other hand, if for some function $m_\psi \in L^2(0, 2\pi)$ satisfying Equation (19) and so defining the function $\psi \in L^2(\mathbb{R})$, the matrix $M(\xi)$ is a.e. unitary, then $\{\psi_{0,k}\}_{k\in\mathbb{Z}}$ isorthonormal in W_0.

Proof. We first note that

$$M^*M = \begin{pmatrix} |m_\varphi(\xi)|^2 + |m_\varphi(\xi+\pi)|^2 & \overline{m_\varphi(\xi)}m_\psi(\xi) + \overline{m_\varphi(\xi+\pi)}m_\psi(\xi+\pi) \\ m_\varphi(\xi)\overline{m_\psi(\xi)} + m_\varphi(\xi+\pi)\overline{m_\psi(\xi+\pi)} & |m_\psi(\xi)|^2 + |m_\psi(\xi+\pi)|^2 \end{pmatrix},$$

and that therefore, we only need to check that the diagonal entries are equal to 1, and that only one of the others are equal to 0 (since they are equal up to conjugation). Therefore, let $\{\psi_{0,k}\}_{k\in\mathbb{Z}}$ be an orthonormal sequence in W_0. We know from Equation (12) that

$$\sum_{k\in\mathbb{Z}} |\widehat{\varphi}(2\xi + 2k\pi)|^2 = 1, \tag{21}$$

for almost every $\xi \in \mathbb{R}$ (since the equality does not depend on ξ except on a null set). Since m_φ is 2π-periodic, and using the filter equation on $\widehat{\varphi}(2\xi + 2k\pi)$, we get that

$$\begin{aligned}\widehat{\varphi}(2\xi + 2k\pi) =& m_\varphi(\xi + k\pi)\widehat{\varphi}(\xi + k\pi) \\ =& \begin{cases} m_\varphi(\xi)\widehat{\varphi}(\xi + 2l\pi) & \text{if } k = 2l, \\ m_\varphi(\xi + \pi)\widehat{\varphi}(\xi + (2l+1)\pi) & \text{if } k = 2l+1 \end{cases}.\end{aligned}$$

Therefore, splitting the sum in (21), we get that

$$\begin{aligned} 1 =& \sum_{l\in\mathbb{Z}} \left|m_\varphi(\xi)\right|^2 |\widehat{\varphi}(\xi + 2l\pi)|^2 \\ &+ \sum_{l\in\mathbb{Z}} \left|m_\varphi(\xi + \pi)\right|^2 |\widehat{\varphi}(\xi + (2l+1)\pi)|^2 \\ =& \left|m_\varphi(\xi)\right|^2 + \left|m_\varphi(\xi + \pi)\right|^2 \sum_{l\in\mathbb{Z}} |\widehat{\varphi}(\xi + 2\pi)| \\ =& \left|m_\varphi(\xi)\right|^2 + \left|m_\varphi(\xi + \pi)\right|^2. \end{aligned}$$

The last step is possible just by noticing that the periodization of an orthonormal system is invariant under translations (since again, the identity holds for a.e. ξ). Showing the same identity but with m_ψ is exactly the same procedure, so we omit that part of the proof.

Now it remains to show that

$$m_\varphi(\xi)\overline{m_\psi(\xi)} + m_\varphi(\xi + \pi)\overline{m_\psi(\xi + \pi)} = 0. \tag{22}$$

From Plancherel, and the fact that $\psi_{0,k} \in W_0$, we have that for any $l, k \in \mathbb{Z}$, $\psi_{0,k} \perp \varphi_{0,l}$

$$0 = \langle \psi(x-k), \varphi(x-l) \rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{\psi}(\xi)\overline{\widehat{\varphi}(\xi)} e^{-i(k-l)\xi}\, d\xi.$$

Setting $n = k - l$, we can rewrite this as a projection in $L^2_p(0, 2\pi)$ onto the basis element $e^{-in\xi}$ we get

$$\begin{aligned}\int_{-\infty}^{\infty} \widehat{\psi}(\xi)\overline{\widehat{\varphi}(\xi)} e^{-in\xi}\, d\xi =& \sum_{k\in\mathbb{Z}} \int_{2k\pi}^{2(k+1)\pi} \widehat{\psi}(\xi)\overline{\widehat{\varphi}(\xi)} e^{-in\xi}\, d\xi \\ =& \sum_{k\in\mathbb{Z}} \int_0^{2\pi} \widehat{\psi}(\xi + 2k\pi)\overline{\widehat{\varphi}(\xi + 2k\pi)} e^{-in\xi}\, d\xi \\ =& \int_0^{2\pi} \left(\sum_{k\in\mathbb{Z}} \widehat{\psi}(\xi + 2k\pi)\overline{\widehat{\varphi}(\xi + 2k\pi)} \right) e^{-in\xi}\, d\xi.\end{aligned}$$

But this means that as a 2π-periodic function, the function inside the parentheses of the last integral is just identically 0 a.e., since it is orthogonal to the basis $\{e^{ikx}\}_{k\in\mathbb{Z}}$. In particular, we have that

$$\sum_{k\in\mathbb{Z}} \widehat{\psi}(2\xi + 2k\pi)\overline{\widehat{\varphi}(2\xi + 2k\pi)} = 0 \text{ for a.e.} \xi \in \mathbb{R}.$$

Therefore, we can redo the steps as in the first part of this proof, and end up with the desired identity. Going through the motions

$$\begin{aligned}\sum_{k\in\mathbb{Z}} \widehat{\psi}(2\xi + 2k\pi)\overline{\widehat{\varphi}(2\xi + 2k\pi)} &= \sum_{l\in\mathbb{Z}} m_\varphi(\xi + 2l\pi)\overline{m_\psi(\xi + 2l\pi)}\,|\widehat{\varphi}(\xi + 2l\pi)|^2 \\ &+ \sum_{l\in\mathbb{Z}} m_\varphi(\xi + (2l+1)\pi)\overline{m_\psi(\xi + (2l+1)\pi)}\,|\widehat{\varphi}(\xi + (2l+1)\pi)|^2 \\ &= m_\varphi(\xi)\overline{m_\psi(\xi)} + m_\varphi(\xi + \pi)\overline{m_\psi(\xi + \pi)} = 0 \text{ for a.e. } \xi \in \mathbb{R}.\end{aligned}$$

This concludes proving (a), but actually, since all the steps are reversible, this actually means we have also proved (b). That is, if for the matrix $M(\xi)$ defined above, it holds almost everywhere that $M^*M = I$, then the sequence $\{\psi_{0,k}\}_{k\in\mathbb{Z}}$ is orthonormal in W_0. ☺

A natural question is how one would go about constructing ψ when all we have is a multiresolution analysis. Further, we don't just want $\{\psi_{0,k}\}_{k\in\mathbb{Z}}$ to just be orthonormal in W_0, we also want it to be a basis, so that we actually end up with a Wavelet basis for all of $L_2(\mathbb{R})$. A partial answer is given by Theorem 42 (b). Note that by Equation (22), we have that

$$\begin{pmatrix} m_\psi(\xi) \\ m_\psi(\xi + \pi) \end{pmatrix} \perp \begin{pmatrix} m_\varphi(\xi) \\ m_\varphi(\xi + \pi) \end{pmatrix} \text{ for a.e. } \xi \in \mathbb{R},$$

as vectors in $\mathbb{C}^2$ over $\mathbb{C}$, with the standard inner product. From linear algebra, we know that in finite dimensions (at least), orthogonality implies linear indepenence. Further, we also know that

$$\begin{pmatrix} a \\ b \end{pmatrix} \perp \begin{pmatrix} \overline{b} \\ -\overline{a} \end{pmatrix}.$$

Since we are in two dimensions, this means that the vector $(\overline{m_\varphi(\xi + \pi)}, -\overline{m_\varphi(\xi)})^T$ is parallel to $(m_\psi(\xi), m_\psi(\xi + \pi))^T$. In other words, there exists some function $\gamma(\xi)$ for $\xi \in \mathbb{R}$ such that

$$\begin{pmatrix} m_\psi(\xi) \\ m_\psi(\xi + \pi) \end{pmatrix} = \gamma(\xi) \begin{pmatrix} \overline{m_\varphi(\xi + \pi)} \\ -\overline{m_\varphi(\xi)} \end{pmatrix} \text{ for a.e. } \xi \in \mathbb{R}.$$

To get more out of this function γ we can note that since both m_ψ and m_φ are 2π-periodic, we also have that

$$\gamma(\xi + \pi) = -\gamma(\xi) \tag{23}$$

almost everywhere, since

$$m_\psi(\xi) = \gamma(\xi)\overline{m_\varphi(\xi + \pi)} = -\gamma(\xi + \pi)\overline{m_\varphi(\xi + \pi)}. \tag{24}$$

Further, since both vectors have length 1 (this follows from the matrix M being unitary), we also have that

$$|\gamma(\xi)| = 1 \text{ a.e..} \tag{25}$$

So if we have a function satisfying (23) and (25), we automatically get a function $\psi \in L^2(\mathbb{R})$ for which $\{\psi_{0,k}\}_{k\in\mathbb{Z}}$ is orthonormal and lives in W_0. It is not hard to check that the function $-e^{-\xi}$ satisfies both equations for all $\xi \in \mathbb{R}$. So take $m_\psi(\xi) = -e^{-i\xi}\overline{m_\varphi(\xi+\pi)}$ from Equality (24) and we have

$$\begin{aligned} m_\psi(\xi) &= -\frac{e^{-i\xi}}{\sqrt{2}}\overline{\sum_{k\in\mathbb{Z}} c_k e^{-ik(\xi+\pi)}} \\ &= -\frac{1}{\sqrt{2}}\sum_{k\in\mathbb{Z}}(-1)^k\overline{c_k}e^{-i\xi(1-k)}. \end{aligned}$$

Making use of the filter identity (Equation (17)), we know that our function ψ satisfies

$$\widehat{\psi}(\xi) = m_\psi(\xi/2)\cdot\widehat{\varphi}(\xi/2) \iff \psi(x) = \overline{\mathscr{F}}\left[m_\psi(\xi/2)\cdot\widehat{\varphi}(\xi/2)\right](x) \text{ for a.e. } x\in\mathbb{R}.$$

Computing the right-hand side of the above gives us that

$$\begin{aligned} \psi(x) =& \frac{1}{2\pi}\int_{\mathbb{R}} e^{i\xi x} m_\psi(\xi/2)\widehat{\varphi}(\xi/2)\,d\xi \\ =& -\frac{1}{2\pi}\int_{\mathbb{R}} e^{i\xi x}\sum_{k\in\mathbb{Z}}(-1)^k\overline{c_k}\frac{1}{\sqrt{2}}e^{-i\frac{\xi}{2}(1-k)}\widehat{\varphi}(\xi/2)\,d\xi \\ =& \sum_{k\in\mathbb{Z}}(-1)^{k+1}\overline{c_k}\frac{1}{\sqrt{2}}\left(\frac{1}{2\pi}\int_{\mathbb{R}} e^{i\frac{\xi}{2}(2x+k-1)}\widehat{\varphi}(\xi/2)\,d\xi\right) \\ . \end{aligned}$$

A change of variables $\xi = 2\eta$ and $1-k = l$ gives us that the last integral is just equal to $\varphi(2x-l)$ a.e., and so we finally get that

$$\psi(x) = \sum_{l\in\mathbb{Z}}(-1)^l\overline{c_{1-l}}\varphi_{1,l}(x) \text{ for a.e. } x\in\mathbb{R}.$$

As stated earlier, the way we have constructed ψ makes it so that sequence $\{\psi_{0,k}\}_{k\in\mathbb{Z}}$ is orthonormal, and is contained in W_0, and is. All that is left now is to show that this sequence also makes up a basis of W_0. This next theorem is essentially the grand finale of this project. It has been a long and arduous journey, but we can now create wavelet bases as we want, given that we have a multiresolution analysis!

Theorem 43. Suppose $\{V_j\}_{j\in\mathbb{Z}}$ is a multiresolution analysis of $L^2(\mathbb{R})$ with φ as its scaling function. If $\psi \in L^2(\mathbb{R})$ is defined by

$$\psi(x) = \sum_{k\in\mathbb{Z}}(-1)^k\overline{c_{1-k}}\varphi_{1,k}(x),$$

where c_k are the structure constants in (16), then $\{\psi_{0,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis of W_0.

Proof. We have already shown that $\{\psi_{0,k}\}_{k\in\mathbb{Z}}$ is orthonormal in W_0, so all that remains to show is that for any $f \in W_0$, there exists a sequence $\{a_k\}_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$ such that

$$f(x) = \sum_{k\in\mathbb{Z}} a_k\psi_{0,k}(x). \tag{26}$$

Since $f \in W_0$, f is orthogonal to V_0, and in particular, it is orthogonal to $\varphi(x)$, and so we can repeat the argument made in Theorem 42 to yield that

$$m_f(\xi) = \gamma(\xi)\overline{m_\varphi(\xi + \pi)}, \tag{27}$$

for some $\gamma \in L_p^2(0, 2\pi)$ satisfying $\gamma(\xi + \pi) = -\gamma(\xi)$ a.e. on $\mathbb{R}$.

Taking the Fourier transform of ψ, we have that

$$\begin{aligned}
\widehat{\psi}(\xi) &= \int_{\mathbb{R}} e^{-i\xi x} \left(\sum_{k\in\mathbb{Z}} (-1)^k \overline{c_{1-k}} \varphi_{1,k}(x) \right) dx \\
&= \sum_{k\in\mathbb{Z}} (-1)^k \overline{c_{1-k}} \widehat{\varphi_{1,k}}(\xi) \\
&= \sum_{k\in\mathbb{Z}} (-1)^k \overline{c_{1-k}} e^{-ik\xi/2} \frac{1}{\sqrt{2}} \widehat{\varphi}(\xi/2) \\
&= \widehat{\varphi}(\xi/2) \frac{1}{\sqrt{2}} \sum_{k\in\mathbb{Z}} \overline{c_{1-k}} e^{-ik(\xi/2+\pi)} \\
&= \widehat{\varphi}(\xi/2) \frac{1}{\sqrt{2}} \sum_{l\in\mathbb{Z}} \overline{c_l} e^{-i(1-l)(\xi/2+\pi)} \\
&= \widehat{\varphi}(\xi/2) e^{-i(\xi/2+\pi)} \overline{m_\varphi(\xi/2 + \pi)}.
\end{aligned}$$

This last equality looks familiar to the one of $\widehat{f}$, and by making making use of (27), we get that

$$\begin{aligned}
\widehat{f}(\xi) &= m_f(\xi/2)\widehat{\varphi}(\xi/2) \\
&= \gamma(\xi/2)\overline{m_\varphi(\xi/2 + \pi)}\widehat{\varphi}(\xi/2) \\
&= \gamma(\xi/2)e^{i(\xi/2+\pi)}\widehat{\psi}(\xi).
\end{aligned}$$

Now let

$$\theta(\xi) = e^{i(\xi/2+\pi)}\gamma(\xi/2), \xi \in \mathbb{R}.$$

If θ lives in $L_p^2(0, 2\pi)$ we are done. We have that

$$\theta(\xi) = -e^{i\xi/2}\beta(\xi/2) = e^{i(\xi/2+2\pi)}\beta(\xi/2 + \pi) = \theta(\xi + 2\pi).$$

Further,

$$\int_0^{2\pi} |\theta(\xi)|^2 \, d\xi = \int_0^{2\pi} \left| \gamma(\frac{\xi}{2}) \right|^2 d\xi = \frac{1}{2} \int_0^{4\pi} |\theta(\xi)|^2 \, d\xi \leq \infty,$$

and so we are done, we have shown that $\theta(\xi)$ is in actuality a filter of f, since we can rewrite θ as a Fourier series with coefficients $\{a_k\}_{k\in\mathbb{Z}} \in \ell(\mathbb{Z})$, and get that

$$\widehat{f}(\xi) = \sum_{k\in\mathbb{Z}} a_k e^{ik\xi}\widehat{\psi}(\xi),$$

which is what we wanted. ☺

Example 44. As promised at the beginning of this section, we are now ready to show that the Haar wavelet introduced back in Example 32 generates a wavelet basis. Notice that

$$\begin{aligned}\psi(x) =&\chi_{[0,1/2)}(x) - \chi_{[1/2,1)}(x) = \frac{\sqrt{2}}{\sqrt{2}}\left(\chi_{[0,1)}(2x) - \chi_{[0,1)}(2x-1)\right)\\ =&\frac{1}{\sqrt{2}}\left(\varphi_{1,0}(x) - \varphi_{1,1}(x)\right) = \sum_{k\in\mathbb{Z}}(-1)^k\overline{c_{1-k}}\varphi_{1,k}(x),\end{aligned}$$

and so ψ is just the wavelet generated by the Haar system, meaning the sequence $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ is an orthonormal basis of $L^2(\mathbb{R})$!

6 Further reading

There is a lot more to be said. The Measure theory covered in this thesis could be expanded to handle things like principal values, a technicality of integration we have chosen to ignore for the purposes of brevity. For a more involved exposition of measure theory, the reader is urged to read [AXL20]. Along those lines, Fourier analysis could also have been explored further, as there are a lot of things that have been left unsaid on the topic. Things like convolutions, the uncertainty principle, and the windowed Fourier transform just to start with, are all deeply connected to the material of this thesis, and helps put Wavelets into context even more. If the reader would like to know more, [GW99] and [TUR18] are sources that I have found to be thoroughly enjoyable to read.

References

[AXL20] SHELDON AXLER, *Measure, integration & real analysis*, Springer, 2020.

[GW99] C. GASQUET and P. WITOMSKI, *Fourier analysis and applications: Filtering, numerical computation, wavelets*, Springer, 1999.

[TUR15] BENGT OVE TURESSON, *An introduction to wavelets*, 2015.

[TUR18] ______, *Fourier analysis, distribution theory, and wavelets*, 2018.

- I Exempel 33 har jag definierat Haar Waveleten fel. Den ska vara 1 på $[0, \frac{1}{2}]$ och -1 på $[\frac{1}{2}, 1]$.