EXAMENSARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

Wavelets - an introduction

av

Karin Gambe

2008 - No 3

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 10691 STOCKHOLM

Wavelets - an introduction

Karin Gambe

Examensarbete i matematik 30 högskolepoäng, fördjupningskurs

Handledare: Rikard Bøgvad

2008

ABSTRACT. The purpose of this paper is to give a brief introduction to wavelets and the wavelet transform. In order to reach this goal the paper starts with some basic properties Hilbert spaces. The wavelet transform has a lot in common with the windowed Fourier transform but while the Fourier window is rigid the wavelet varies depending on the time/frequency ratio. In applications wavelets are often used together with a multiresolution analysis (MRA) and towards the end it will be shown how a wavelet basis is constructed from a MRA.

Contents

1. Introduction

The main purpose of this paper is to define wavelets. In brief a wavelet can be described as a "‘helping"’ function that is used to extract information specific in time and frequency from a more complex function.
Let us say that you have a sound signal with a lot of buzz or unwanted noise. If you use wavelets you can divide the signal into wavelet coefficients with a decomposition algorithm, throw away the coefficients corresponding to the unwanted frequencies, put the modified signal back together with the reconstruction algorithm and achieve a much cleaner sound, without loosing any vital information.
To be able to do this, this so-called wavelet has to be both time- and frequency oriented. Here comes wavelets' big advantage in application like this to other similar algorithms: their ability to zoom in on both time and frequency so that they are able to adjust to sudden irregularities.
One definition of wavelets is that a wavelet is a function with zero mean

$$\int_{-\infty}^{\infty} \psi(x)dx = 0$$

which can be scaled by a parameter a and translated by another parameter b

$$\psi_{b:a}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)$$

so that it by translation and scaling can cover the whole time-frequency plane. A wavelet is only allowed to exist for a short period of time, like a wave appearing from nowhere on the ocean, only to grow over a reef and then to break and disappear. This is also where the name comes from. It is an English modification of the french word "‘ondelette"’, meaning small wave. It was first introduced by the french geophysicist Jean Morlet in the 1970's. He needed a function that was localized in both time and frequency and that could adjust its time interval to sudden peaks in the seismic signals. He found a group of functions constructed by Alfred Haar in the beginning of the 20^{th} century and started to use them for his seismic data. This was the start of the development in wavelet theory which today is applied in a wide range of areas, from solving differential equations to compressing data.

To be able to define wavelets and describe some of their applications, this paper will start with a detour in the world of functional analysis. We need to get familiar with Hilbert spaces in general and the space $L^2(\mathbb{R})$ in particular since that is the space where most of the theory of wavelets take place.
Then some approximation theories are described, mostly in order to motivate the need of wavelets, but also to describe Fourier analysis

which is vital in the wavelet theory.
Then it is time to describe window functions in general only to move on and describe the wavelet transforms. With the chapter of multiresolution analysis the construction of wavelets are implemented and with the decomposition and reconstruction some use of multiresolution and wavelets are described.

2. The space $L^2(\mathbb{R})$

In this first section we introduce the space $L^2(\mathbb{R})$ and some of its important properties. This is since the wavelets studied in this paper are defined in this space. The space $L^2(\mathbb{R})$ is the normed space of square-integrable functions, and to be able to understand what that means some essential concepts in functional analysis will be listed. For a total review of these concepts, see [7].

Definition 2.1. ***Norm***
A norm is a real-valued function on linear space X such that, for $x \in X$ it takes x to $\|x\|$. The norm $\|x\|$ of x has the following properties

- $\|x\| \geq 0$
- $\|x\| = 0 \Leftrightarrow x = 0$
- $\|\alpha x\| = |\alpha| \, \|x\|$
- $\|x + y\| \leq \|x\| + \|y\|$

The norm gives a metric on X defined by $d(x, y) = \|x - y\|$.

Definition 2.2. ***Normed space***
A normed space X is a vector space with a norm.

Definition 2.3. ***Complete space***
A metric space X is complete if every Cauchy sequence in X converges to a point in X.

Definition 2.4. ***Banach space***
A Banach space is a complete normed space. Note that it has to be complete in the metric defined by the norm.

Example 2.5. *ℓ^p is the space of all sequences $x = (x_1, x_2, ..)$ such that $\sum |x_i|^p < \infty$.*
ℓ^p has the norm

$$\|x\| = (\sum |x_i|^p)^{\frac{1}{p}} \tag{2.1}$$

and the metric

$$d(x, y) = (\sum |x_i - y_i|^p)^{1/p} \tag{2.2}$$

so the space is a normed space. To see whether or not it is a Banach space, let us check if any Cauchy sequence in ℓ^p converges.

Let $\{x_n\}$ be any Cauchy sequence in ℓ^p.
Then for every ϵ there exists an N_ϵ so that for all $m, n > N$

$$d(x_m, x_n) = (\sum |x_i^{(m)} - x_i^{(n)}|^p)^{1/p} < \epsilon. \tag{2.3}$$

For this to hold $|x_i^{(m)} - x_i^{(n)}|$ must be smaller than $< \epsilon$ for every $i = 1, 2..$, so let us take a fixed i and consider the sequence $(x_i^{(1)}, x_i^{(2)}, ..)$. It is a Cauchy sequence of numbers and it converges for both real and complex numbers since both $\mathbb{R}$ and $\mathbb{C}$ are complete metric spaces.
Let $x_i^{(m)} \to x_i$ as $m \to \infty$.
Put $x = (x_1, x_2, ..)$. Let $n \to \infty$ in (2.3), then for $m > N$

$$\sum_{i=1}^{k} |x_i^{(m)} - x_i|^p \leq \epsilon^p \tag{2.4}$$

Preceding and letting $k \to \infty$, for $m > N$

$$\sum_{i=1}^{\infty} |x_i^{(m)} - x_i|^p \leq \epsilon^p \tag{2.5}$$

which is the same as $(d(x_m - x))^p \leq \epsilon^p$. So $x_m \to x$. It is necessary that x is in ℓ^p, but (2.5) shows that $x_m - x = (x_i^{(m)} - x_i) \in \ell^p$ and since $x_m \in \ell^p$ the Minkowski inequality (later introduced as theorem 2.19) says that x also is in ℓ^p.
Since (x_n) was an arbitrary Cauchy sequence in ℓ^p this example has showed that the space is complete, hence it is a Banach space.

Example 2.6. *Denote by $\mathbb{Q}$,the set of all rational numbers . It has a norm $\|x\| = |x|$ that induces the metric $d(x, y) = |x - y|$. So $\mathbb{Q}$ is a normed space but it is not complete and hence not a Banach space.*

Definition 2.7. ***Inner product, inner product space***
An inner product on a vector space X is a function that to each pair (x, y) associates a number denoted $\langle x, y\rangle \in X$ that satisfies

- $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$
- $\langle \alpha x, y\rangle = \alpha \langle x, y\rangle$
- $\langle x, y\rangle = \overline{\langle y, x\rangle}$. *(If X is a real space then $\langle x, y\rangle = \langle y, x\rangle$)*
- $\langle x, x\rangle \geq 0$, $\langle x, x\rangle = 0 \Leftrightarrow x = 0$.

An inner product space is a vector space X with an inner product defined on X.

The inner product defines a norm on X

$$\|x\| = \sqrt{\langle x, x\rangle} \tag{2.6}$$

and a metric

$$d(x, y) = \|x - y\| = \sqrt{\langle x - y, x - y\rangle}. \tag{2.7}$$

Hence it is possible to conclude that spaces with inner products are normed spaces. For the converse to be true it is necessary that the norm can be obtained from an inner product, so observe that not all normed spaces are inner product spaces.

Theorem 2.8. *Let $\|*\|$ be any norm on a vector space X. Then the following are equivalent:*

(1) *$\|*\|$ is induced by a unique inner product on X.*
(2) *$\|a+b\|^2 + \|a-b\|^2 = 2(\|a\|^2 + \|b\|^2)$ for all $a,b \in X$.*

The first part of this proof is straightforward. The second part will require some additional information in form of a lemma which will be left unproven. (See[5].) The proof of the theorem itself illustrates many properties of inner product spaces.

Lemma 2.9. *Let $\|*\|$ be any norm on a vector space X over the field K such that $\|a+b\|^2 + \|a-b\|^2 = 2(\|a\|^2 + \|b\|^2)$, for all $a,b \in X$. Define a function $F : X^2 \to \mathbb{R}$ as*

$$F(x,y) = \frac{1}{4}(\|x+y\|^2 + \|x-y\|^2) \quad \textit{for all} \quad (x,y) \in X^2$$

Then, for all $x, y, z \in X$ and $\lambda \in K$

(1) *$F(x+y,z) = F(x,z) + F(y,z)$;*
(2) *$8(F(\lambda x, y) + F(x, \lambda y)) = (|\lambda+1|^2 - |\lambda-1|^2)(\|x+y\|^2 + \|x-y\|^2)$.*

Proof. (Of theorem 2.8)

(1) Let $\|*\|$ be a norm induced by an inner product, then

$$\|a+b\|^2 + \|a-b\|^2 = \langle a+b, a+b\rangle + \langle a-b, a-b\rangle$$
$$= \|a\|^2 + \|b\|^2 + \langle a,b\rangle + \langle b,a\rangle + \|a\|^2 + \|b\|^2 - \langle a,b\rangle - \langle b,a\rangle$$
$$= 2(\|a\|^2 + \|b\|^2).$$

(2) Let F be defined as in Lemma 2.9 and begin to consider the case $K = \mathbb{R}$. Let us check that F is an inner product and that the norm is induced by F. Take $x, y, z \in X$ and $\alpha \in \mathbb{R}$. By definition of F it is possible to check that $F(x,y) = F(y,x)$ and $F(x,x) = \|x\|^2 \geq 0$, with the equality only when $x = 0$. The first part in Lemma shows that F is bilinear so that

$$F(\alpha x, y) = \alpha F(x,y) \tag{2.8}$$

Hence it is an inner product on X and the norm is induced by F.

Let $K = \mathbb{C}$, let F be as above (note that it is realvalued). Define a function G: $G(x,y) = F(x,y) + iF(x,iy)$.
By Lemma 2 and by equation (2.8)

$$G(x+z,y) = G(x,y) + G(z,y)$$
$$G(\alpha x, y) = \alpha G(x,y) \text{ for every } \alpha \in \mathbb{R}.$$

Use the second part of Lemma 2 and set $\lambda = i$. Then

$$F(ix,y) = -F(x,iy) \tag{2.9}$$

and

$$F(ix, iy) = F(x, y) \tag{2.10}$$

so that

$$F(ix, y) + iF(ix, iy) = iF(x, y) - F(x, iy)$$

and

$$G(ix, y) = iG(x, y).$$

Let $t = \alpha + \beta i$ for $(\alpha, \beta) \in \mathbb{R}^2$. Then

$$G(tx, y) = G(\alpha x, y) + G(\beta ix, y)$$

$$= \alpha G(x, y) + \beta G(ix, y) = tG(x, y).$$

Recall that F is real valued, so that

$$F(u, v) = \overline{F(v, u)} = F(v, u) \text{ for all } u, v \in X.$$

Together with (2.9) this gives

$$G(y, x) = F(y, x) + iF(y, ix) = \overline{F(x, y)} - \overline{iF(x, iy)} = \overline{G(x, y)}.$$

Since $F(x, ix) = 0$ then

$$G(x, x) = F(x, x) + iF(x, ix) = \|x\|^2$$

so that $G(x, x) \geq 0$ with equality only if $x = 0$. This shows that G is an inner product and that the norm $\|*\|$ is induced by G.

The uniqueness follows from the fact that a norm on a vector space only can be induced by at most one inner product. □

The second part of Theorem 2.8 is called the parallellogram law and it gives an easy way to check whether a norm can be obtained from an inner product or not.

Example 2.10. *As seen in the previous example, the space ℓ^p is a normed space, but for $p \neq 2$ it is not an inner product space.*
To check this, let us use Theorem 2.8 and set

$$x = (1, 1, 0, 0, ...) \text{ and } y = (1, -1, 0, 0, ...)$$

then $x, y \in \ell^p$ but

$$\|x\| = \|y\| = 2^{1/p} \neq \|x + y\| = \|x - y\| = 2 \text{ when } p \neq 2. \tag{2.11}$$

The parallelogram equality is not satisfied when $p \neq 2$ which shows that the space ℓ^p with $p \neq 2$ is not an inner product space.

Definition 2.11. ***Hilbert space***
A Hilbert space is a complete inner product space with the norm defined by the inner product.

Example 2.12. *As we checked above the space ℓ^2 is both an inner product space and complete, hence it is a Hilbert space. It is in fact the prototype Hilbert spaces. Hilbert himself used it in his work.*

The following theorem states that any normed space can be completed into an Hilbert space. There is a similar theorem regarding Banach spaces which can be found in [7] where also the proof of this theorem can be found. This theorem is important since the space $L^2(\mathbb{R})$ is the completion of a normed space.

Theorem 2.13. *For any inner product space X there exists a Hilbert space H and an isomorphism A from X onto a dense subspace W of H. H is unique up to isomorphism.*

Example 2.14. *The space $L^2[a,b]$ has the norm*

$$\|x\| = (\int_a^b |x(t)|^2 dt)^{1/2} \tag{2.12}$$

which can be obtained from the inner product

$$\langle x, y\rangle = \int_a^b x(t)\overline{y(t)}dt. \tag{2.13}$$

So it is an inner product space, and since $L^2[a,b]$ is complete with this norm (as the completion of the space X with the same norm) it is also a Hilbert space.

Definition 2.15. *$L^2(\mathbb{R})$*
The space $L^2(\mathbb{R})$ is the completion of the vector space of all continuous functions on $\mathbb{R}$.

Using inner product in an inner product space it is possible to define the notion of orthogonality. This concept will be used applied to wavelet bases.

Definition 2.16. ***Orthogonality***
Let X be an inner product space, then for $x, y \in X$, we say that they are orthogonal if

$$\langle x, y\rangle = 0.$$

(The orthogonality is denoted $x \perp y$.)

Example 2.17. *Let $X = L^2[0,1]$ and let*

$$\phi(x) = \begin{cases} 1 & \text{if } 0 \le x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.14}$$

and

$$\psi(x) = \begin{cases} 1 & \text{if } 0 \le x < 1/2 \\ -1 & \text{if } 1/2 \le x < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.15}$$

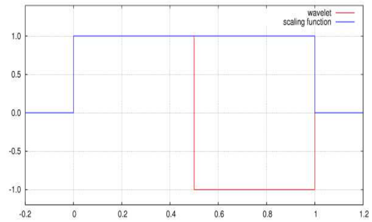

FIGURE 1. The Haar wavelet and its scaling function.

Then ϕ and ψ are orthogonal in $L^2[0,1]$ since

$$\langle \phi, \psi \rangle = \int_0^{1/2} 1dx - \int_{1/2}^1 1dx = 0.$$

We will get back to these functions since, as we will see in section 8, ψ is the wavelet function and ϕ is the corresponding scaling function for the Haar system. (See also Figure 1.)

2.1. **Properties of $L^2(\mathbb{R})$.** In the remaining part of this section some important properties of Hilbert spaces will be listed. Their proofs can be found in [7].

Theorem 2.18. *(Hölder inequality for series)*
For $p > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$

$$\sum_{j=1}^{\infty} |x_j y_j| \leq \left(\sum_{k=1}^{\infty} |x_k|^p \right)^{1/p} \left(\sum_{m=1}^{\infty} |y_m|^q \right)^{1/q}. \tag{2.16}$$

*For $p = 2$ this is the **Cauchy-Schwartz Theorem**.*

Theorem 2.19. *(Minkowski inequality for sums)*
Let $x, y \in \ell^p$ and $p \geq 1$, then

$$\left(\sum_{j=1}^{\infty} |x_j - y_j| \right)^{1/p} \leq \left(\sum_{k=1}^{\infty} |x_k|^p \right)^{1/p} \left(\sum_{m=1}^{\infty} |y_m|^p \right)^{1/p}. \tag{2.17}$$

Theorem 2.20. *(Schwartz inequality)*
Let $\langle x, y \rangle$ be an inner product and let $\|*\|$ be its corresponding norm. Then

$$|\langle x, y \rangle| \leq \|x\| \, \|y\| \, . \tag{2.18}$$

Equality if and only if $\{x, y\}$ is linearly independent.

Theorem 2.21. *(Continuity of inner product)*
Let X be an inner product space and let $x, y \in X$.
If $x_n \to x$ and $y_n \to y$ as $n \to \infty$ then

$$\langle \, x_n, y_n \rangle \to \langle \, x, y \rangle \, . \tag{2.19}$$

The proof of this result is so elegant and short that it is included.

Proof.

$$\begin{aligned}
&|\langle\, x_n, y_n\rangle - \langle\, x, y\rangle| \\
&= |\langle\, x_n, y_n\rangle - \langle\, x_n, y\rangle + \langle\, x, y_n\rangle - \langle\, x, y\rangle| \\
&\leq |\langle\, x_n, y_n - y\rangle| + |\langle\, x_n - x, y\rangle| \\
&\leq \|x_n\| \, \|y_n - y\| + \|x_n - x\| \, \|y\| \to 0 \text{ as } n \to \infty.
\end{aligned} \tag{2.20}$$

since $x_n \to x$ and $y_n \to y$ as $n \to \infty$

□

2.2. **Direct sum.** The theory of direct sums is used for the multiresolution analysis and for the decomposition algorithm.
The aim of the first part of this subsection is to provide the definition of the direct sum. For proofs and further explanation of the theorems and definitions, see [7].

Theorem 2.22. *Let X be a complete metric space. Then a subspace Y of X is complete if and only if Y is closed in X.*

Definition 2.23. ***The distance to a subspace.***
Let X be a metric space, then the distance δ from X to a nonempty subset Y is given by

$$\delta = \inf_{\tilde{y} \in Y} d(x, \tilde{y}). \tag{2.21}$$

If X is a normed space, equation (2.21) coincides with

$$\delta = \inf_{\tilde{y} \in Y} \|x - \tilde{y}\|.$$

The following theorem states that for subspaces which fulfill certain criteria, there exists an unique point y in a subspace which is closest to the point x in the space X. For a proof of this existence and uniqueness problem, see Theorem 3.3-1 in [7].

Theorem 2.24. *Let X be an inner product space. Let Y be a nonempty, convex subset which is complete. Then for every $x \in X$ there exists a unique $y \in Y$ such that*

$$\delta = \inf_{\tilde{y} \in Y} \|x - \tilde{y}\| = \|x - y\|.$$

Lemma 2.25. *Let Y be as in the previous theorem.*
Take $x \in X$. Then $z = x - y$ is orthonormal to Y.

Definition 2.26. ***Direct sum***
Let X be a vector space. Then X is said to be the direct sum of the subspaces Y and Z of X if each $x \in X$ has the unique representation

$$x = y + z$$

The direct sum is then denoted by

$$X = Y \oplus Z \tag{2.22}$$

and Y, Z is said to be a complementary pair of subspaces in X.

Example 2.27. *Let ϕ be the function in (2.14), where*

$$\phi_{j,k} = 2^{j/2}\phi(2^j x - k).$$

Then $\{\phi_{j,k} : j,k \in \mathbb{Z}\}$ generate the space V_j. If $\phi(x) \in V_j$, then $\phi(2x) \in V_{j+1}$, and if $\phi(x) \in V_j$ then $\phi(x + 2^{-j}) \in V_j$. Hence the sequence

$$...V_{j-1} \subset V_j \subset V_{j+1}$$

is a nested sequence. The space V_j is a proper subspace of V_{j+1}, which means that $V_j \neq V_{j+1}$. Let us call this quotent space W_j. It is generated by $\{\psi_{j,k} : j,k \in \mathbb{Z}\}$, where $\psi_{j,k} = 2^{j/2}\psi(2^j x - k)$. The W_j-spaces are mutually orthogonal and

$$V_j \cap W_j = \{0\}$$

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

Theorem 2.28. *Let H be a Hilbert space and let Y be any closed subspace of H. Then*

$$H = Y \oplus Y^{\perp} \tag{2.23}$$

where $Y^{\perp} = \{x \in H \mid x \perp Y\}$.

Proof. Since H is complete and Y is closed then Y is complete.
There exists a $y \in Y$ for every $x \in H$ such that

$$x = y + z \ , \ z \in Y^{\perp}.$$

The uniqueness is proved as follows. Assume that

$$x = y + z = y' + z' \text{ where } y, y' \in X \text{ and } z, z' \in Y^{\perp}$$

Then

$$y - y' \in Y \text{ and } z' - z \in Y^{\perp}$$

so

$$y - y' = z' - z.$$

Y and $Y^{\perp}$ are orthogonal, so $Y \cap Y^{\perp} = \{0\}$. Since $y - y' \in Y \cap Y^{\perp}$, $y - y' = 0$ and hence $y = y'$. The same procedure is used to show that $z = z'$. □

3. Frames and Riesz basis.

A non-precise definition of a wavelet is that a wavelet is a function ψ such that $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$, with

$$\psi_{j,k} = 2^{j/2}\psi(2^j x - k)$$

forms an orthonormal basis of $L^2(\mathbb{R})$, which also can be relaxed so that $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ only has to constitute a Riesz basis of $L^2(\mathbb{R})$. The aim of this section is to understand how those two definitions correspond to each other but also to introduce the concepts of frames and Riesz basis.

3.1. **Orthonormal and dual basis.**

Definition 3.1. ***Orthonormal basis***
A basis $\{e_j\}$ is an orthonormal basis if $\|e_j\| = 1$ for all j and if

$$\langle e_i, e_j \rangle = \delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases} \tag{3.1}$$

The symbol δ is called the Kronecker delta.

Dual bases are needed for example to define R-wavelets.

Definition 3.2. ***Dual basis***
Let X be a vector space and $\{e_1, e_2, ..., e_n\}$ be a basis of X. Then the set of all linear functionals on X constitutes the algebraic dual space X^* of X. For every functional f and every $x = \sum x_i e_i \in X$ such a functional can be written as

$$f(x) = \sum x_i f(e_i)$$

Every set $f(e_1), ..., f(e_n)$ determines a linear functional on X, so with

$$f_j(e_i) = \delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

there will be n functionals denoted by $f_1, f_2, ..., f_n$. The basis $\{f_1, f_2, ..., f_n\}$ is called the dual basis to $\{e_1, e_2, ..., e_n\}$ in X.

Definition 3.3. ***Unconditional basis***
If $\sum_n \mu_n e_n \in X$ implies that $\sum_n |\mu_n| e_n \in X$, then the basis $\{e_n\}_{n\in\mathbb{Z}}$ is said to be unconditional.

In an unconditional basis the convergence of a series with this basis does not depend on the order of summation of entries in this series.

3.2. **Frames.**

The theory of frames is needed for the discrete form of wavelet transforms and for dyadic wavelets.
Frames can be thought of as a more general form of a basis. Vectors that constitutes a frame spans a Hilbert space H but they do not have to be linearly independent. If a function can be represented by a frame, then the function has a stable representation.

Definition 3.4. ***Frames***
Let H be a Hilbert space and let $0 < A \leq B < \infty$ be positive constants. Then $\psi = \{\psi_j : j \in J\}$is said to generate a frame of H if

$$A \|f\|^2 \leq \sum_{j \in J} |\langle f, \psi_j \rangle|^2 \leq B \|f\|^2 \text{ for all } f \in H. \tag{3.2}$$

A and B are called the frame bounds.

If H is the space $L^2(\mathbb{R})$ then the summation is made for $j, k \in \mathbb{Z}$.
If $A = B$ the frame is called tight and then (3.2) gives

$$\sum_{j \in J} |\langle f, \psi_j \rangle|^2 = A \|f\|^2 .$$

Given some extra assumptions a tight frame constitutes an orthonormal basis.

Proposition 3.5. *Let $\{\psi_j : j \in J\}$ be a tight frame and let $A = B = 1$. If $\|\psi_j\| = 1$ for all $j \in J$, then ψ_j generates an orthonormal basis.*

Proof. If $\langle f, \psi_j \rangle = 0$ for all f, then $f = 0$ by the properties of the inner product, so ψ_j span H. For any $j \in J$

$$\|\psi_j\|^2 = \sum_{j' \in J} |\langle \psi_j, \psi_{j'} \rangle|^2 = \|\psi_j\|^4 + \sum_{j' \neq j \in J} |\langle \psi_j, \psi_{j'} \rangle|^2.$$

Since $\|\psi_j\| = 1$ the $\langle \psi_j, \psi_{j'} \rangle$ has to be 0 for all $j' \neq j$ which makes $\{\psi_j : j \in J\}$ orthonormal. $\square$

To show that extra conditions on $\|\psi_j\|$ and A, B are necessary in order to get an orthonormal basis, lets consider an example.

Example 3.6. *Let $H = \mathbb{C}^2$ and set*
$e_1 = (0, 1)$ $e_2 = \left(-\frac{\sqrt{3}}{2}, -\frac{1}{2}\right)$ *and* $e_3 = \left(\frac{\sqrt{3}}{2}, -\frac{1}{2}\right)$.
Then e_1, e_2, e_3 are not linearly independent but for any $v = (v_1, v_2) \in H$ we get

$$\begin{aligned}
&\textstyle\sum_j |\langle v, e_j \rangle|^2 \\
&= |v_2|^2 + \left|-\tfrac{\sqrt{3}}{2} v_1 - \tfrac{1}{2} v_2\right|^2 + \left|\tfrac{\sqrt{3}}{2} v_1 + \tfrac{1}{2} v_2\right|^2 \\
&= \tfrac{3}{2} \left(|v_1|^2 + |v_2|^2\right) = \tfrac{3}{2} \|v\|^2
\end{aligned}$$

which shows that $\{e_1, e_2, e_3\}$ is a tight frame with frame bounds $A = \frac{3}{2}$. But since the e_j's are not linearly independent they can not give an orthonormal basis.

For further use the notion of a dual frame will be needed. The easiest way to define the dual frame is to use the 'frame operator'.

Definition 3.7. ***Frame operator***
Let $\{\psi_j : j \in J\}$ be a frame of H. Then the linear operator F from H to $\ell^2(J)$ defined by

$$(Ff)(j) = \langle f, \psi_j \rangle$$

is called the frame operator.

Definition 3.8. ***Adjoint operator***
Let $F : H_1 \to H_2$ be a bounded linear operator from one Hilbert space to another. Then the adjoint operator F^* is the operator

$$F^* : H_2 \to H_1$$

such that for all $x \in H_1$ and all $y \in H_2$

$$\langle Fx, y \rangle = \langle x, F^* y \rangle .$$

Remark: F is bounded since by definition (3.7)

$$\|Ff\|^2 = B\,\|f\|^2 .$$

By theorem 3.9-2 in [7] the adjoint operator F^* of F always exists and is unique and it is a bounded linear operator with the same norm as F. That is

$$\|F^*\| = \|F\| .$$

Moreover, for every positive bounded linear operator F in a Hilbert space which is bounded from below by C, there exists an inversion F^{-1} of F such that F^{-1} is bounded by C^{-1} from above (see [7]).

By $A \leq \|FF^*\| \leq B$ it is obvious that (FF^*) is bounded below and hence it is invertible.

Definition 3.9. ***Dual frame***
The dual frame $\tilde{\psi}_j$ of ψ_j is defined by $\tilde{\psi}_j = (FF^*)^{-1}(\psi_j)$ and

$$B^{-1}\,\|f\|^2 \leq \sum_{j\in J} |\left\langle f, \tilde{\psi}_j \right\rangle|^2 \leq A^{-1}\,\|f\|^2 \text{ for all } f \in H. \tag{3.3}$$

3.3. **Riesz basis.** The notion af a Riesz basis is crucial to describe wavelets. A Riesz basis is a stronger formulation of a frame but is still weaker than an orthogonal basis. Even though orthogonal wavelets are the most convinient it is often enough to use wavelets which constitute a Riesz basis only.

Definition 3.10. ***Riesz basis***
A set $\{\beta_{j,k}\}$ is called a Riesz basis if

(1) *the linear span $\langle \beta_{j,k} : j, k \in \mathbb{Z}\rangle$ is dense in $L^2(\mathbb{R})$, and*
(2)

$$A\,\|\{c_{j,k}\}\|_{\ell^2}^2 \leq \left\| \sum_{j,k\in\mathbb{Z}} c_{j,k}\beta_{j,k} \right\|^2 \leq B\,\|\{c_{j,k}\}\|_{\ell^2}^2 \tag{3.4}$$

for all $c_{j,k} \in \ell^2$ and for $0 < A \leq B < \infty$, where A and B are called Riesz bounds.

As is the case of frames in Proposition 3.5; if one has $A = B = 1$, then the Riesz basis is orthonormal.
A Riesz basis can also be defined as an unconditional basis in a Hilbert space.

Example 3.11. *The function $\phi(t) = \frac{\sin \pi t}{\pi t}$ is called the* ***Shannon scaling function****. The sequence $\{\phi_{j,k}\}$ spans a family of the nested spaces V_j given in example 2.27. A Shannon basis is an orthonormal Riesz basis.*

If A is a bounded operator with a bounded inverse, then A maps any orthonormal basis to a Riesz's basis. That is, if a Hilbert space H is finite dimensional, every basis in H is a Riesz basis. Since our space $L^2(\mathbb{R})$ is not finite, this does not apply to that space.

Definition 3.12. ***R-function***
Let β be a function which satisfies equation (3.4)*. Then β is called an R-function.*

As seen from their definitions frames and Riesz basis have a lot in common. The following theorem clarifies the difference between them.

Theorem 3.13. *Let $\psi \in L^2(\mathbb{R})$, then the following statements are equivalent:*

- **:** *(i) $\{\psi_{j,k}\}$ is a Riesz basis of $L^2(\mathbb{R})$*
- **:** *(ii) $\{\psi_{j,k}\}$ is a frame of $L^2(\mathbb{R})$ which is also an linearly independent family in ℓ^2.*

The frame bounds and the Riesz bounds do agree.

Proof. $(i) \Rightarrow (ii)$:
Because of (3.4), any Riesz basis is ℓ^2-linearly independent.
Let $\{\psi_{j,k}\}$ be a Riesz basis with bounds A and B.
Let $M = [\gamma_{l,m:j,k}]_{(l,m)(j,k)\in\mathbb{Z}^2}$ be a linear operator where

$$\gamma_{l,m:j,k} = \langle \psi_{l,m}, \psi_{j,k} \rangle . \tag{3.5}$$

Put M in (3.4)

$$A \left\| \{c_{j,k}\} \right\|_{\ell^2}^2 \leq \left\| \sum_{j,k\in\mathbb{Z}} c_{l,m}\gamma_{l,m:j,k}\bar{c}_{j,k} \right\|^2 \leq B \left\| \{c_{j,k}\} \right\|_{\ell^2}^2$$

to show that M is positive definite. Then

$$M^{-1} = [\rho_{l,m:j,k}]_{(l,m)(j,k)\in\mathbb{Z}^2}$$

is the inverse of M and

$$\sum_{r,s} \rho_{l,m:r,s}\gamma_{r,s:j,k} = \delta_{l,j}\delta_{m,k} \tag{3.6}$$

and

$$B^{-1} \left\| \{c_{j,k}\} \right\|_{\ell^2}^2 \leq \left\| \sum_{j,k\in\mathbb{Z}} c_{l,m}\rho_{l,m:j,k}\bar{c}_{j,k} \right\|^2 \leq A^{-1} \left\| \{c_{j,k}\} \right\|_{\ell^2}^2 \tag{3.7}$$

holds, so it is possible to write

$$\psi^{l,m}(x) = \sum \rho_{l,m:j,k}\psi_{j,k}(x).$$

Since $\psi^{l,m}$ is in $L^2(\mathbb{R})$ then (3.5) and (3.6) give that

$$\left\langle \psi^{l,m}\psi_{j,k} \right\rangle = \delta_{l,j}\delta_{m,k}.$$

Thus $\{\psi^{l,m}\}$ is the dual basis to $\{\psi_{j,k}\}$ in $L^2(\mathbb{R})$. Now (3.6) and (3.7) gives that

$$\left\langle \psi^{l,m}, \psi^{j,k} \right\rangle = \rho_{l,m:j,k}$$

so that the Riesz bounds of $\{\psi^{l,m}\}$ are A^{-1} and B^{-1}.
Any $f \in L^2(\mathbb{R})$ can be written

$$f(x) = \sum \langle f, \psi_{j,k} \rangle \, \psi^{j,k}(x)$$

and

$$B^{-1} \sum | \langle f, \psi_{j,k} \rangle |^2 \leq \|f\|^2 \leq A^{-1} \sum | \langle f, \psi_{j,k} \rangle |^2$$

which is the same as (3.3).

For the implication $(ii) \Rightarrow (i)$, see the proof of Theorem 3.20 in [3]. $\square$

The conclution of this theorem is that a Riesz basis is a frame of linearly independent vectors.

4. Approximation theories

Approximation can be used in a huge area of applications. One is to compress or filter data with the help of wavelets. To do it we need to find a wavelet that can be used to represent a more complicated function. Also in other application of approximation the goal is to represent complicated functions by simpler functions. Once this is achieved, the function can be represented by this simple function and after some modifications hopefully there exists a way to reconstruct the original data as close to the original as desired.
Approximation theory is a big field but the aim of this section is to motivate the need of wavelets.

4.1. **Weierstrass approximation.**

The Weierstrass theorem says that any continuous function on a closed and bounded interval can be approximated by a polynomial. More exactly:

Theorem 4.1. *Let f be a continuous function on a closed and bounded interval $I \in \mathbb{R}$. Then, for any $\epsilon > 0$ there exists a polynomial P such that*

$$|f(x) - P(x)| \leq \epsilon \text{ for all } x \in I.$$

The Weierstrass theorem only states the existence of a polynomial P, but it is still interesting that no matter how small ϵ is, there will always exist some P corresponding to it.

4.2. **Power series.** The following theorem is one form of the Taylor theorem that can be found in any undergraduate textbook on calculus. (See for example [6].)

Theorem 4.2. *Let f be a smooth function defined on an interval I. If there exists $C > 0$ such that $|f^{(n)}(x)| \leq C$ for all $n \in \mathbb{N}$ and for all $x \in I$, then for $x_0 \in I$*

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n \text{ for all } x \in I.$$

Remark: It is convenient that it is so easy to find the approxikating polynomial even though it limits the set of functions that can be approximated pretty hard. The power series can only represent functions that are smooth.

5. Fourier Analysis

With the approximation theory described so far, only really well-behaved functions can be approximated.
Using Fourier analysis it is possible to approximate a larger class of functions by means of trigonometric functions. Fourier analysis describes the spectral behavior of a function, if there is a function in a time-frequency plane for example, then it is convenient to use Fourier analysis to find out what happens on the frequency axis. Since Fourier analysis is a necessary tool for wavelets, its role in this paper is not only as another version of approximation theory. As a general reference for this section, see [1] and [3].

5.1. **Fourier series.** Fourier series is often used to describe the behavior of discrete functions. It describes a trigonometric expansion of a function $f(x)$ in a series.
The expansion of f will be on the form

$$F(x) = a_0 + \sum a_n \cos(nx) + b_n \sin(nx). \tag{5.1}$$

Here f is 2π periodic so that $f(x+2\pi) = f(x)$ for $x \in \mathbb{R}$ and $f \in L^2(\mathbb{R})$.
Since

$$\int_{\pi+c}^{\pi+c} f(x)dx = \int_{\pi}^{\pi} f(x)dx \tag{5.2}$$

it does not matter what the interval looks like as long as its length is 2π.

Theorem 5.1. *Let $f(x) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos(\frac{n\pi x}{a}) + b_n \sin(\frac{n\pi x}{a})\right)$. Then for $x \in (-a, a)$ one has*

$$a_0 = \frac{1}{2a}\int_{-a}^{a} f(t)dt \tag{5.3}$$

and

$$a_n = \frac{1}{a}\int_{-a}^{a} f(t)\cos\left(\frac{n\pi t}{a}\right)dt \tag{5.4}$$

$$b_n = \frac{1}{a}\int_{-a}^{a} f(t)\sin\left(\frac{n\pi t}{a}\right)dt \tag{5.5}$$

This is a generalization of the normal form with $x = \frac{t\pi}{a}$ *and* $dx = \frac{\pi dt}{a}$.

If $S = f(x)$, then $S_N(x) = a_0 + \sum_{n=1}^{N}(a_n \cos(\frac{n\pi x}{a}) + b_n \sin(\frac{n\pi x}{a})$ is the partial sum of $f(x)$.

Example 5.2. *Set*

$$f(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Let $a = 2$ *so that the Fourier series of* f *is valid on the interval* $[-2, 2]$. *Then*

$$a_0 = \frac{1}{4}\int_{-2}^{2} f(t)dt = \frac{1}{4}\int_0^1 1dt = \frac{1}{4}$$

and for $n \ge 1$

$$a_n = \frac{1}{2}\int_{-2}^{2} f(t)\cos\left(\frac{n\pi t}{2}\right)dt\frac{1}{a}\int_0^1 \cos\left(\frac{n\pi t}{2}\right)dt = \frac{\sin(n\pi/2)}{n\pi}.$$

If n *is even,* $a_n = 0$ *and if* n *is odd then* $a_n = \sin\left(\frac{n\pi}{2}\right) = (-1)^k$ *so that*

$$a_n = \frac{(-1)^k}{(2k+1)\pi}\ ,\ n = 2k+1.$$

$$b_n = \frac{1}{2}\int_{-2}^{2} f(t)\sin\left(\frac{n\pi t}{2}\right)dt = \frac{1}{2}\int_0^1 \sin\left(\frac{n\pi t}{2}\right)dt = \frac{-1}{n\pi}(\cos(\frac{n\pi}{2})-1)$$

Totally

$$n = 4j \Rightarrow b_n = 0$$
$$n = 4j+1 \Rightarrow b_n = \frac{1}{(4j+1)\pi}$$
$$n = 4j+2 \Rightarrow b_n = \frac{1}{(2j+1)\pi}$$
$$n = 4j+3 \Rightarrow b_n = \frac{1}{(4j+3)\pi}$$

Hence the Fourier series of f *can be written as in (5.1).*

An even function is a function $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) = f(-x)$. An odd function is a function such that $f(-x) = -f(x)$. Examples of even functions are x^2 or $\cos(x)$ and odd functions could be x^3 or $\sin(x)$. For even functions one has

$$\int_{-a}^{a} f(x)dx = 2\int_0^a f(x)dx, \text{ for all } a > 0 \tag{5.6}$$

and for odd functions

$$\int_{-a}^{a} f(x)dx = 0, \text{ for all } a > 0. \tag{5.7}$$

The above properties gives some simple consequences for their Fourier series since $\cos(x)$ is even and $\sin(x)$ is odd and since for functions

$$\text{Odd} \times \text{Odd} = \text{Even}$$
$$\text{Even} \times \text{Even} = \text{Even}$$
$$\text{Even} \times \text{Odd} = \text{Odd}.$$

This gives the following

- The Fourier series of even functions can only contain cosines since $b_n = 0$ for all n so that

$$f(x) = a_0 + \sum a_n \cos\left(\frac{n\pi x}{a}\right) \text{ and } a_n = \frac{2}{a}\int_0^a f(x)\cos\left(\frac{n\pi x}{a}\right) dx$$

- Analoguesly: the Fourier series of odd functions only contain sines since $a_n = 0$ for all n so that

$$f(x) = \sum b_n \sin\left(\frac{n\pi x}{a}\right) \text{ and } b_n = \frac{2}{a}\int_0^a f(x)\sin\left(\frac{n\pi x}{a}\right) dx.$$

Example 5.3. *Let $f(x) = x$ for $x \in [-\pi, \pi]$.*
Since f is an odd function, then only the sine-coefficients are to be computed. One has

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} x\sin(nx)dx = \frac{2(-1)^{k+1}}{k},$$

which gives the Fourier series expansion

$$F(x) = \sum \frac{2(-1)^{k+1}}{k}\sin(kx).$$

Fourier series approximate the original function well as long as it is continuous.
In the points where the function jumps (is discontinuous) the Fourier series will overshoot. This is called the **Gibb's phenomena** (see for example [1]). If f is approximated by S_N the overshooting will get smaller as N gets bigger.
To achieve a pointwise match between the Fourier series and the function f it is necessary to assume that f is continuous .

Definition 5.4. ***Piecewise smooth***
Let f be a continuous function. Then f is called piecewise smooth if its derivatives are defined everywhere except for on a discrete set of points.

Lemma 5.5. *Let $f(x)$ be as in (5.1) and let $\sum_{n=1}^{\infty}(|a_n| + |b_n|) < \infty$. Then the Fourier series of $f(x)$ converges uniformly and absolutely to $f(x)$.*

This lemma will be used to show the following result.

Theorem 5.6. *Let $f(x)$ be a piecewise smooth, and 2π-periodic function. Then its Fourier series converges uniformly to $f(x)$ on $[-\pi,\pi]$ and*

$$|f(x) - S_N(x)| \leq \frac{1}{\sqrt{N}}\frac{1}{\sqrt{\pi}}\sqrt{\int_{-\pi}^{\pi}|f'(t)|^2dt}.$$

Proof. To make life easier, assume that f is twice continuously differentiable. Let $f(x) = \sum(a_n\cos(nx)+b_n\sin(x))$ and $f''(x) = \sum(a_n''\cos(nx)+b_n''\sin(x))$, then $a_n = -\frac{a_n''}{n^2}$ and $b_n = -\frac{b_n''}{n^2}$ since

$$\begin{aligned} a_n &= \tfrac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)dx \\ &= \left[f(x)\tfrac{\sin(nx)}{n}\right]_{-\pi}^{\pi} - \tfrac{1}{\pi}\int_{-\pi}^{\pi} f'(x)\tfrac{\sin(nx)}{n}dx \\ &= 0 - b_n' = \tfrac{-1}{n^2\pi}\int_{-\pi}^{\pi} f''(x)\cos(nx)dx = -\tfrac{a_n''}{n^2} \end{aligned} \tag{5.8}$$

b_n'' is derived in the same manner. Since f'' is continuous, the Riemann-Lebesgue theorem gives that a_n'' and b_n'' converges to zero as $n \to \infty$. This means that

$$\sum_{n=1}^{\infty}|a_n| + |b_n| = \sum_{n=1}^{\infty}\frac{|a_n''|+|b_n''|}{n^2} \leq \sum_{n=1}^{\infty}\frac{M+M}{n^2} < \infty$$

and the proof follows from Lemma 5.5. □

The next theorem describes how to get a value of the Fourier transform at points where f is discontinuous.

Theorem 5.7. *Let f be a piecewise continuous and 2π-periodic function. Let $F(x)$ be its Fourier series.*

- *If f is continuous at a point x, then $F(x)$ converges and*
$$F(x) = f(x).$$
- *If f is discontinuous at x_j but is left- and right-differentiable at x_j, then*
$$F(x) = \frac{1}{2}\left(\lim_{x\to x_j^+} f(x) + \lim_{x\to x_j^-} f(x)\right)$$

A proof of both parts of the theorem can be found in [1], Theorems 1.22 and 1.28.

Example 5.8. *Let $f(x)$ be the 2π-periodic extension of $y = x$, $-\pi \leq x \leq \pi$ as in Example 5.3.Then f is discontinuous at $x = ... - \pi, \pi...$, and the left and right limits at $x = \pi$ are equal*

$$f(\pi - 0) = \lim_{x\to\pi^+} f(x) = \pi$$

$$f(\pi + 0) = \lim_{x\to\pi^-} f(x) = -\pi.$$

Since the derivatives

$$f'(\pi - 0) = 1$$

and

$$f'(\pi + 0) = 1$$

exist, f is left- and right-differentiable at $x = \pi$.
The average of the limits are $\frac{\pi+(-\pi)}{2} = 0$, so that $F(\pi k) = 0$.
Consider the expansion $F(x) = \sum \frac{2(-1)^{k+1}}{k} \sin(kx)$ from Example 5.3.
It is also zero for $x = \pi$.

The following proposition shows that the Fourier series for functions in $L^2[-\pi, \pi]$ converges almost everywhere. This also implies that any function in $L^2[-\pi, \pi]$ can be approximated arbitrarily close by a smooth 2π-periodic function.

Proposition 5.9. *Let $f \in L^2[-\pi, \pi]$ be a continuous, piecewise differentiable and 2π-periodic function. Let a_n and b_n be the Fourier coefficients of f. Then,*

$$|f(x) - S_N(x)| \leq \sum_{n=N+1}^{\infty} (|a_n| + |b_n|) \text{ for all } x \in \mathbb{R}.$$

Proof. By Theorem 5.7 the assumptions implies that the Fourier series tends to $f(x)$ for all $x \in \mathbb{R}$. Hence

$$\begin{aligned}
&|f(x) - S_N(x)| \leq \\
&\left| a_0 + \sum_{n=1}^{\infty} (a_n \cos(nx) + b_n \sin(nx)) - \left(a_0 + \sum_{n=1}^{N} (a_n \cos(nx) + b_n \sin(nx)) \right) \right| \\
&= \left| \sum_{n=N+1}^{\infty} (a_n \cos(nx) + b_n \sin(nx)) \right| \\
&\leq \sum_{n=N+1}^{\infty} |a_n \cos(nx) + b_n \sin(nx)| \\
&\leq \sum_{n=N+1}^{\infty} (|a_n| + |b_n|).
\end{aligned}$$

□

Even though Fourier series are a very useful tool when approximating periodic functions it is not useful for represention of non-periodic functions.

5.2. **Fourier transform.** Fourier transform can be thought of as a continuous version of Fourier series. It describes the spectral behavior of continuous functions.
Even though the focus has and will be on the space $L^2(\mathbb{R})$, this section start with Fourier transform in the space $L^1(\mathbb{R})$ simply to make things easier.

Definition 5.10. *For $f \in L^1(\mathbb{R})$ its Fourier transform is defined by $\widehat{f}(\omega) = (Ff)(\omega) = \int_{-\infty}^{\infty} e^{-i\omega x} f(x) dx$.*

Example 5.11. *Let*

$$f(x) = \begin{cases} 1 & if \ -\pi \le x \le \pi \\ 0 & otherwise \end{cases} \tag{5.9}$$

$$f(x)e^{-\lambda x} = f(x)(\cos \lambda x - i \sin \lambda x).$$

Since f is even, so $f(x)\sin\lambda x$ becomes odd and the integral of $f(x)\sin\lambda x$ over the real line vanishes. Thus

$$\hat{f}(\lambda) = \frac{1}{\sqrt{2\pi}} \int f(x)\cos(\lambda x)dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \cos(\lambda x)dx = \frac{\sqrt{2}\sin(\lambda\pi)}{\sqrt{\pi}\lambda}.$$

The Fourier transform has the following properties.

Theorem 5.12. *Let $f \in L^1(\mathbb{R})$. Then the Fourier transform $\hat{f}$ of f satisfies*

(1) $\hat{f} \in L^\infty$ *and* $\left\|\hat{f}\right\|_\infty \le \|f\|_1$
(2) $\hat{f}$ *is uniformly continuous on* $\mathbb{R}$
(3) *if f' exists and is in $L^1(\mathbb{R})$, then*

$$\hat{f'}(\omega) = i\omega \hat{f}(\omega) \tag{5.10}$$

(4) $\hat{f}(\omega) \to 0$ *as* $\omega \to \pm\infty$.

It is not clear that $\hat{f}(\omega) \in L^1(\mathbb{R})$ just because $\hat{f}$ tends to zero as $\omega \to \pm\infty$, as shown by the following example.

Example 5.13. *Let $f(x) = e^{-x}u_0(x)$ with*

$$u_0(x) = \begin{cases} 1 & if\ x \ge a \\ 0 & if\ x < a \end{cases}$$

Then $f(x) \in L^1(\mathbb{R})$ but,

$$\hat{f}(x) = \textstyle\int_0^\infty e^{-x} cos(\omega x)dx - i \int_0^\infty e^{-x} sin(\omega x)dx$$
$$= \tfrac{1}{1+\omega^2} - \tfrac{i\omega}{1+\omega^2} = \tfrac{1}{1-\omega}$$

Since $\frac{1}{1-\omega}$ is not in $L^1(\mathbb{R})$, neither is $\hat{f}(x)$.

However, we will soon see that when both f and its Fourier transform are in $L^1(\mathbb{R})$ it is possible to reconstruct, or recover, f from $\hat{f}$.

Definition 5.14. ***Inverse Fourier transform***
Let $\hat{f} \in L^1(\mathbb{R})$ be the Fourier transform of a function $f \in L^1(\mathbb{R})$. Then

$$(F^{-1}\hat{f})(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega x} \hat{f}(\omega) \tag{5.11}$$

is the inverse Fourier transform of $\hat{f}$.

The interesting part will be to check when $(F^{-1}\hat{f}) = f$. We prove it in the inverse Fourier series theorem below. To be able to state that theorem we will need some definitions.

Definition 5.15. ***Convolution***
Let $f, g \in L^1(\mathbb{R})$ then the convolution $f * g$ is given by

$$(f * g)(x) = \int_{-\infty}^{\infty} f(x-y)g(y)dy. \tag{5.12}$$

The convolution has the property that when $f, g \in L^p(\mathbb{R})$, the convolution $f * g$ is in $L^p(\mathbb{R})$ as well.

Definition 5.16. ***Gaussian function***
The function on the form $f(x) = e^{-x^2}$ is called Gaussian. A special family of Gaussian functions are defined as

$$g_\alpha(t) = \frac{1}{2\sqrt{\pi\alpha}} e^{\frac{t^2}{4\alpha}} \tag{5.13}$$

With convolution the Gaussian functions has the property that if $f \in L^1(\mathbb{R})$ and f is continuous in x, then

$$(f * g_\alpha)(x) \to f(x) \text{ as } \alpha \to 0^+.$$

For the proof of the following theorem about invertibility of the Fourier transform the identity

$$\int_{-\infty}^{\infty} f(x)\hat{g}(x) = \int_{-\infty}^{\infty} \hat{f}(x)g(x) \tag{5.14}$$

for functions $f, g \in L^1(\mathbb{R})$ will be needed.
It holds since $\hat{f}, \hat{g} \in L^\infty(\mathbb{R})$ by Theorem 5.12 and the integrals are finite by Hölder inequality.

Theorem 5.17. *Take $f, \hat{f} \in L^1(\mathbb{R})$ and let f be continuous in x. Then*

$$f(x) = \left(F^{-}\hat{f}\right)(x).$$

Proof. Fix an $x \in L^1(\mathbb{R})$ and set

$$g(y) = \frac{1}{2\pi} e^{iyx} e^{-\alpha y^2}.$$

Then

$$\begin{aligned} \hat{g}(y) &= \tfrac{1}{2\pi} \textstyle\int e^{-iyt} e^{itx} e^{-\alpha t^2} dt \\ &= \tfrac{1}{2\pi} \textstyle\int e^{-i(y-x)t} e^{-\alpha t^2} dt \\ &= \tfrac{1}{2\pi} \sqrt{\tfrac{\pi}{\alpha}} e^{-\frac{(y-x)^2}{4\alpha}} = g_\alpha(x-y), \end{aligned}$$

where g_α is of the form (5.13).
From Theorem 5.12 we know that $\hat{f}, \hat{g} \in L^\infty(\mathbb{R})$ so that the identity

(5.14) holds. Then the convolution in (5.12) can be used together with the **Fubini theorem** (see [9]) so that

$$
\begin{aligned}
&(f * g)(x) = \int f(y) g_{\alpha}(x-y) dy \\
&= \int f(y)\hat{g}(y) dy = \int \hat{f}(y) g(y) dy \\
&= \frac{1}{2\pi} \int e^{iyx} \hat{f}(y) e^{-\alpha y^2} dy.
\end{aligned}
$$

Since f is continuous in x, the right-hand side tends to $(F^{-}\hat{f})(x)$ and the left-hand side converges to $f(x)$ as $\alpha \to 0^{+}$. ☐

By this theorem it is clear that the inverse of the Fourier transform in $L^1(\mathbb{R})$ exist only of the points where f is continuous.

The Fourier transform in $L^1(\mathbb{R})$ satisfies the following properties.

$$
\begin{array}{ll}
\text{Function} & \text{Fourier transform} \\
f(t) & \hat{f}(\omega) \\
f * g(t) & \hat{f}(\omega)\overline{\hat{g}(\omega)} \\
fg(t) & \frac{1}{2\pi}\hat{f} * \hat{g}(\omega) \\
f(t-u) & e^{-i\omega u}\hat{f}(\omega) \\
f(t/s) & |s|\hat{f}(s\omega) \\
\bar{f}(t) & \overline{\hat{f}(-\omega)}
\end{array}
\tag{5.15}
$$

Even though the possibility of extension of the Fourier transform to $L^2(\mathbb{R})$ has not yet been proved, it can be mentioned that all the properties above holds for the Fourier transform in that space as well.

5.3. **Fourier transform in $L^2(\mathbb{R})$.** In $L^1(\mathbb{R})$ the function f has to be continuous in order to get the inverse. In $L^2(\mathbb{R})$, the Fourier transform $\hat{f}$ is a one-to-one and onto mapping which means that it maps $L^2(\mathbb{R})$ to itself which makes it easy to find the inverse $(F^{-1}f)$.
The following theorem has two parts where the first part is known as the **Parseval theorem** and the second as the **Plancherel theorem**. This theorem makes it clear that the Fourier transform can be extended to $L^2(\mathbb{R})$.

Theorem 5.18. (1) *If $f, g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ then*

$$
\int f(t)\overline{g(t)} dt = \frac{1}{2\pi} \int \hat{f}(\omega)\overline{\hat{g}(\omega) d\omega} \ \text{or} \ \langle f, g \rangle = \frac{1}{2\pi} \left\langle \hat{f}, \hat{g} \right\rangle \tag{5.16}
$$

(2)

$$
\int |f(t)|^2 dt = \frac{1}{2\pi} \int \left| \hat{f}(\omega) \right|^2 d\omega \ \text{or} \ \left\| \hat{f} \right\|_2 = 2\pi \left\| f \right\|_2. \tag{5.17}
$$

Proof. Let $h = f * g$ as in (5.12).
One property of the Fourier transform is that $\hat{h}(\omega) = \hat{f}(\omega)\overline{\hat{g}(\omega)}$.

Use the Fourier inverse formula (5.11) with $h(0)$.

$$\int f(t)\overline{g(t)}dt = h(0) = \frac{1}{2\pi}\int \hat{h}(\omega)d\omega = \frac{1}{2\pi}\int \hat{f}(\omega)\overline{\hat{g}(\omega)}d\omega.$$

The second part of the theorem follows by simply letting $g = f$. $\square$

Consider a function $f \in L^2(\mathbb{R})$, such that $f \notin L^1(\mathbb{R})$ since $f(t)e^{i\omega t}$ is not integrable it is not possible to calculate the Fourier transform of f with the Fourier integral.
Instead it is necessary to consider functions $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$.
$L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ is dense in $L^2(\mathbb{R})$ so let $\{f_n\}_{n\in\mathbb{Z}}$ be a family of functions in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, then there must exists a function f such that

$$\|f - f_n\| \to 0 \text{ as } n \to \infty \tag{5.18}$$

(5.18) implies that $\{f_n\}_{n\in\mathbb{Z}}$ is Cauchy, so that for $n, m > N$

$$\|f_n - f_m\| < \epsilon.$$

$\hat{f}$ is well-defined since $f \in L^1(\mathbb{R})$, so by the means of Theorem 5.18 it is possible to show that $\left\{\hat{f}_n\right\}_{n\in\mathbb{Z}}$ is also Cauchy:

$$\left\|\hat{f}_n - \hat{f}_m\right\| = \sqrt{2\pi}\,\|f_n - f_m\| \leq \sqrt{2\pi}\epsilon \text{ when } n, m > N. \tag{5.19}$$

$L^2(\mathbb{R})$ is complete so there must exist $\hat{f} \in L^2(\mathbb{R})$ such that

$$\left\|\hat{f} - \hat{f}_n\right\| \to 0 \text{ as } n \to \infty. \tag{5.20}$$

This shows that $\hat{f}$ is the Fourier transform of f in $L^2(\mathbb{R})$.

Next we prove the invertibility of the Fourier transform in $L^2(\mathbb{R})$. As mentioned before the inversion in $L^2(\mathbb{R})$ will be much easier to find. Another good thing is that in $L^2(\mathbb{R})$ all nice properties of Hilbert spaces can be applied.
Also, note that the identity (5.14) for functions in $L^1(\mathbb{R})$ also applies to functions in $L^2(\mathbb{R})$

Definition 5.19. ***Reflection of*** f
For every f defined on $\mathbb{R}$, the reflection f^- of f relative to the origin is defined as

$$f^-(x) = f(-x). \tag{5.21}$$

The reflection has the following property

$$\overline{\hat{f}}(x) = (\widehat{\overline{f}^-})(x)\ ;\ (\widehat{\overline{f}^-})(x) = (\hat{f})^-(x). \tag{5.22}$$

Theorem 5.20. *For every $g \in L^2(\mathbb{R})$ there is one and only one $f \in L^2(\mathbb{R})$ such that $\hat{f} = g$. This means that*

$$f(x) = (F^{-1}g)(x) = \check{g}(x) \tag{5.23}$$

is the inverse Fourier transform of g.

Proof. This proof will use equations (5.14) and (5.22). By (5.14) we see that if $g \in L^2(\mathbb{R})$ then $\hat{g} \in L^2(\mathbb{R})$. The following calculations will show that the function $f(x) = \frac{1}{2\pi}\widehat{(g^-)}$ is in $L^2(\mathbb{R})$ and that it satisfy the relation $\hat{f} = g$.

$$\begin{aligned} \left\|g - \hat{f}\right\|_2^2 &= \|g\|_2^2 - 2Re\left\langle g, \hat{f}\right\rangle + \left\|\hat{f}\right\|_2^2 \\ &= \|g\|_2^2 - 2Re\left\langle \hat{g}, f^-\right\rangle + \left\|\hat{f}\right\|_2^2 \\ &= \|g\|_2^2 - 2Re\left\langle \hat{g}, \tfrac{1}{2\pi}\hat{g}\right\rangle + \left\|\hat{f}\right\|_2^2 \\ &= \tfrac{1}{2\pi}\|\hat{g}\|_2^2 - \tfrac{2}{2\pi}\|\hat{g}\|_2^2 + 2\pi\|f\|_2^2 \\ &= -\tfrac{1}{2\pi}\|\hat{g}\|_2^2 + \tfrac{1}{2\pi}\left\|\widehat{g^-}\right\|_2^2 = 0. \end{aligned} \tag{5.24}$$

That is $\hat{f} = g$ as required.
Uniqueness:
If f is the only function in $L^2(\mathbb{R})$ that satisfies $\hat{f} = g$ then $\hat{f} = 0 \Rightarrow f = 0$.
Using the Parseval Theorem, assume $\hat{f} = 0$. Then

$$\langle f, f\rangle = \frac{1}{2\pi}\left\langle \hat{f}, \hat{f}\right\rangle = 0$$

and

$$\langle f, f\rangle = 0 \Leftrightarrow f = 0.$$

□

At the end of the section let us state the Poisson theorem. The Poisson sum is used when constructing wavelets and the proof gives a nice relation between Fourier series and the Fourier transform.

Theorem 5.21. *Let both the series*

$$\sum_{n=-\infty}^{\infty} f(t + 2\pi n)$$

and

$$\sum_{k=-\infty}^{\infty} \hat{f}(k)e^{ikt}$$

be convergent. Then

$$\sum_{n=-\infty}^{\infty} f(t + 2\pi n) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} \hat{f}(k)e^{ikt} \tag{5.25}$$

Proof. Let $f(t) \in L^2(\mathbb{R})$ and let

$$f_p(t) = \sum_{n=-\infty}^{\infty} f(t + 2\pi n)$$

be the 2π-periodic version of $f(t)$. Let the period be 2π, then the Fourier series of $f(t)$ becomes

$$f_p(t) = \sum_{k=-\infty}^{\infty} c_k e^{ikt}$$

with

$$c_k = \frac{1}{2\pi}\int_0^{2\pi} f_p(t)e^{-ikt}dt = \frac{1}{2\pi}\int_0^{2\pi}\sum_{n\in\mathbb{Z}} f(t+2\pi n)e^{-ikt}dt$$

Put $x = t + 2\pi n$. Then

$$\frac{1}{2\pi}\sum_{n\in\mathbb{Z}}\int_{2\pi n}^{2\pi(n+1)} f(x)e^{-ik(x-2\pi n)}dx.$$

This can be extended to get

$$c_k = \frac{1}{2\pi}\int f(x)e^{-ikt}dx = \frac{1}{2\pi}\hat{f}(k)$$

which gives the Poisson formula. □

6. Windowed transforms

6.1. **Windowed Fourier Transform.** To be able to analyze such functions as sound signals it is necessary to localize events in time or frequency. To do this a window function is used. A schematic description of a window function is like a rectangle in the time-frequency plane (or the x-y plane). Sometimes it is enough to have a window that is limited on the time (t-, or x-)axis, this is called a time window. A function f whose Fourier transform $\hat{f}$ is limited in the frequency (ω-,or y-)axis is called a frequency window.

Example 6.1. *Let ϕ be a time window function and $f(t)$ be a function that we want to analyze at the interval $[t-b, t+b]$. Then*

$$f(t)\phi(t-b) = f_b(t) \tag{6.1}$$

gives us information about f close to $t = b$. By changing b it is possible to get parts of information from the entire t-axis. To get information regarding the frequency we take the Fourier transform $\hat{\phi}$ of ϕ and use the interval $[\omega - b, \omega + b]$.

To qualify as a window function $\phi(t)$ has to vanish, or *decay*, sufficiently fast. Hence for a time-window $\phi \in L^2(\mathbb{R})$ the following property is common to use:

$$t\phi(t) \in L^2(\mathbb{R}) \tag{6.2}$$

and for a frequency-window the Fourier transform $\hat{\phi(\omega)}$ of $\phi(t)$ should satisfy

$$\omega\hat{\phi}(\omega) \in L^2(\mathbb{R}). \tag{6.3}$$

When both (6.2) and (6.3) are satisfied simultaneously, the function ϕ can be used as a time-frequency window.

Example 6.2. ***The Gabor transform***
The Gabor transform uses a Gaussian function, as in (5.13) to localize the information from the Fourier transform.

Let $f \in L^2(\mathbb{R})$. Then the Gabor transform of f is defined by

$$(G_b^\alpha f)(\omega) = \int_{-\infty}^{\infty} (e^{-i\omega x} f(t)) g_\alpha(t-b) db \tag{6.4}$$

Gaussian functions have the nice properties that the Fourier transform of it is again a Gaussian and that both g_α and $\hat{g_\alpha}$ satisfies the window-function property in (6.2) and (6.3), so it can be used both for time and frequency analysis.
Using a Gabor transform helps us to localize the Fourier transform around $t = b$, and by changing b it can cover the whole time-axis.
We will see later that the Gaussian function is the window function of minimal area.

Example 6.3. *The Haar function in (2.15) can be used as a time window function.*
However, by Theorems 5.12 and inverse FT, since ψ is not continuous, $\hat{\psi}$ is not in $L^1(\mathbb{R})$ and then $\omega\hat{\psi}(\omega)$ can not be in $L^2(\mathbb{R})$. Hence the Fourier transform $\hat{\psi}(\omega)$ does not satisfy the window property, which means that the function is no good as frequency window.

For a function to be able to serve as a window function it is necessary that it has a radius and a center.

Definition 6.4. ***Center***
For a time window function ϕ the center t^* is given by

$$t^* = \frac{1}{\|\phi\|^2} \int_{-\infty}^{\infty} t\, |\phi(t)|^2 dt. \tag{6.5}$$

Definition 6.5. ***Radius***
The radius Δ_ϕ for a window function ϕ is given by

$$\Delta_\phi = \frac{1}{\|\phi\|} \left(\int_{-\infty}^{\infty} (t - t^2)\, |\phi(t)|^2 \right)^{1/2} \tag{6.6}$$

For a frequency window, the frequency center ω^* and radius $\Delta_{\hat{\phi}}$ is defined in the same way, using $\hat{\phi}$ instead of ϕ.

If a function is a time-window centered in t^* and has the radius b, then the window will be the interval $[t^* - b, t^* + b]$ as in example 6.1, and the width of the function will be $2b$.

Example 6.6. *For a Gaussian function, with $\alpha > 0$, the radius of the function is given by*

$$\Delta_{g_\alpha} = \sqrt{\alpha} \tag{6.7}$$

That means that the width of the window is $2\sqrt{\alpha}$

For any time-frequency window the time-frequency area is given by $\Delta_\phi \Delta_{\hat{\phi}}$.
The following "Uncertainty principle" provides that $\Delta_\phi \Delta_{\hat{\phi}} \geq \frac{1}{2}$.

Theorem 6.7. *For a function $\phi \in L^2(\mathbb{R})$ such that both ϕ and $\hat{\phi}$ satisfy the condition (6.2) the inequality*

$$\Delta_\phi \Delta_{\hat{\phi}} \geq \frac{1}{2} \tag{6.8}$$

holds. Equality will occur if and only if ϕ is a Gaussian function. That is $\phi(t) = ce^{iat} g_\alpha(t-b)$, where $c \neq 0, \alpha > 0$ and $a, b \in \mathbb{R}$.

The proof of the theorem can be found in [3]. The theorem says that the only window function that can achieve minimal area is a Gaussian and, as will be seen later, a Gaussian function can not be a wavelet. Since the smaller the area is the more precise the time-frequency localization will be it is important to search for wavelets with as close to minimal area as possible.

When applying a window function to a Fourier transform the result will be a windowed Fourier transform.

Definition 6.8. ***Windowed Fourier transform***
Let $\phi \in L^2(\mathbb{R})$ be a time-frequency window. Then

$$(Sf)(b, \omega) = \int_{-\infty}^{\infty} (e^{-i\omega x} f(t))\phi(t-b)db = \langle f, \phi_{b,\omega} \rangle \tag{6.9}$$

is the windowed Fourier transform of $f(t)$ at the point (b, ω).

Definition 6.9. ***STFT***
Let ϕ be a function such that both ϕ and $\hat{\phi}$ fulfill the criterium for window functions. Then the window Fourier transform with ϕ as its window function is called a "Short-time Fourier transform" or STFT.

The windowed Fourier transform with $\phi = g_\alpha$ will be the Gabor transform, which is an STFT. In addition, the Gabor transform is the STFT with the smallest window, since it is the only function that can give the equality in the uncertainty principle (theorem 6.7).
Depending on what the goal is the optimal size of the window can differ. Sometimes a bigger window than the Gaussian one is required, and then can for example any B-spline be used, see below.

Example 6.10. *For $m = 1$, the B-spline is defined as*

$$N_1(x) = \begin{cases} 1 & \text{when } 0 \le x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

It is not continuous, so the situation is the same as for the Haar function. Actually N_1 is the function

$$\phi(x) = \begin{cases} 1 & \text{if } 0 \le x < 1 \\ 0 & \text{otherwise .} \end{cases} \tag{6.10}$$

in (2.14).
The m^{th}-order B-spline is denoted

$$N_m(x) = \int_0^1 N_{m-1}(x-t)dt \ , \ m \ge 2$$

Its Fourier transform is given by

$$\hat{N}_m(\omega) = \left(\frac{1-e^{-i\omega}}{i\omega}\right)^m = e^{-\frac{im\omega}{2}} \left(\frac{sin(\omega/2)}{\omega/2}\right)^m .$$

N_m has compact support so it is a time window and $\omega\hat{N}_m(\omega) \in L^2(\mathbb{R})$, so N_m can be used for an STFT.

By changing b and ω it is possible to cover the whole time-frequency plane. When the plane is covered the function f can be recovered from the windowed Fourier transform.

Theorem 6.11. *Let $f \in L^2(\mathbb{R})$ then,*

$$f(t) = \frac{1}{2\pi}\int\int (Sf)(b,\omega)\phi(t-b)e^{i\omega t}d\omega db \tag{6.11}$$

and

$$\int |f(t)|^2 dt = \frac{1}{2\pi}\int\int |(Sf)(b,\omega)|^2 \, d\omega db. \tag{6.12}$$

Proof. We use Theorem 5.18.
Start with observation that

$$(Sf)(b,\omega) = e^{-ib\omega}\int f(t)\phi(t-b)e^{i\omega}(b-t)dt = e^{-ib\omega} f * \phi_\omega(b)$$

where $\phi_\omega(t) = \phi(t)e^{i\omega t}$.
Let $(Sf)(b,\omega) = f_\omega(b)$, then

$$\hat{f}_\omega(x) = \hat{f}(x+\omega)\hat{\phi}_\omega(x+\omega) = \hat{f}(x+\omega)\hat{\phi}(x).$$

The Fourier transform of $\phi(t-b)$ is $\hat{\phi}(x)e^{-it\omega}$, so

$$\begin{aligned} &\tfrac{1}{2\pi}\left(\int\int (Sf)(b,\omega)\phi(t-b)e^{i\omega t}dbd\omega\right) \\ &= \tfrac{1}{2\pi}\int\left(\tfrac{1}{2\pi}\int \hat{f}(x+\omega)|\hat{\phi}(x)|^2 e^{it(x+\omega)}dx\right)d\omega \end{aligned}$$

When $f \in L^2(\mathbb{R})$, the Fubini theorem (again, see [9]) gives that the integration can be reversed. The inverse Fourier transform gives

$$\frac{1}{2\pi}\int \hat{f}(x+\omega)e^{it(x+\omega)}d\omega = f(t).$$

Since $\frac{1}{2\pi}\int |\hat{\phi}(x)|^2 d\omega = 1$ the formula (6.11) is proved. The Fourier transform of $(Sf)(b,\omega)$ is $\hat{f}(x+\omega)\hat{\phi}(x)$.
Apply Plancherel to the right-hand side of (6.12):

$$\frac{1}{2\pi}\int\int |(Sf)(b,\omega)|^2 db d\omega = \frac{1}{2\pi}\int \frac{1}{2\pi}\int |\hat{f}(x+\omega)\hat{\phi}(x)|^2 dx d\omega.$$

If the Fubini theorem and Plancherel is used again the following implies formula (6.12).

$$\frac{1}{2\pi}\int |\hat{f}(x+\omega)|^2 d\omega = \|f\|^2$$

□

7. Wavelet transform

Remember the story of Jean Morlet and his seismic signals from the introduction. For him it was necessary to have a window that could vary in its time-radius. The frequency of a function stands in direct proportion to the length of the time interval and to get as good information as possible when using a window function it is desirable that the time interval shortens for high and widens for low frequencies. If you use a windowed Fourier transform you have to re-calculate a new window every time the time-frequency ratio changes and that makes the calculations time demanding and heavy. What Morlet found in his search for a function that could adapt to diffent ratios was the first type of wavelet constructed by Haar. A wavelet is a function ψ that satisfy the window conditions both for ψ and $\hat{\psi}$. It also has zero average, i.e.

$$\int_{-\infty}^{\infty} \psi(x)dx = 0 \tag{7.1}$$

which makes it possible to use a dilation parameter that scales the function so that the corresponding window can zoom in or out for different frequencies.

Example 7.1. *A Gaussian function does not have zero average. Consider for example*

$$\int_{-\infty}^{\infty} e^{-ax^2}dx = \sqrt{\left(\frac{\pi}{a}\right)}.$$

For a Gabor transform the area is $4\Delta_{g_\alpha}\Delta_{\hat{g_\alpha}} = 2$ and the window has the Cartesian product

$$\left[b-\sqrt{\alpha}, b+\sqrt{\alpha}\right] \times \left[\omega - \frac{1}{2\sqrt{\alpha}}, \omega + \frac{1}{2\sqrt{\alpha}}\right]. \tag{7.2}$$

The Gaussian window is good for very high frequencies, but it can not change its ratio in order to analyze low frequencies.
For a wavelet the area still will be $4\Delta_\psi\Delta_{\hat{\psi}}$ but the Cartesian product is

$$[b+at^* - a\Delta_\psi, b+at^* + a\Delta_\psi] \times \left[\frac{\omega^*}{a} - \frac{1}{a}\Delta_{\hat{\psi}}, \frac{\omega^*}{a} + \frac{1}{a}\Delta_{\hat{\psi}}\right], \tag{7.3}$$

so that when the scaling parameter a gets bigger, the window narrows. It can be worth to note that the ratio between the frequency center ω^* and the frequency width $\Delta_{\hat{\psi}}$ is independent of the location of the frequency center which is given by a, i.e.

$$\frac{\omega^*/a}{2\Delta_{\hat{\psi}}/a} = \frac{\omega^*}{2\Delta_{\hat{\psi}}}.$$

With $\frac{\omega^*}{a} = \omega$ the window will adapt to the frequency so that it widens for low and narrows for high frequencies.

Example 7.2. *The Mexican hat is a well known wavelet. It is the second derivative of a Gaussian function and is used a lot in geophysics. The normalized form of the Mexican wavelet is*

$$\psi(t) = \frac{2}{\pi^{1/4}\sqrt{3\sigma}}\left(\frac{t^2}{\sigma^2} - 1\right)e^{\left(-\frac{t^2}{s\sigma^2}\right)} \tag{7.4}$$

The Fourier transform of ψ is

$$\hat{\psi}(\omega) = -\frac{\sqrt{8}\sigma^{5/2}\pi^{1/4}}{\sqrt{3}}\omega^2 e^{\left(-\frac{\sigma^2\omega^2}{2}\right)} \tag{7.5}$$

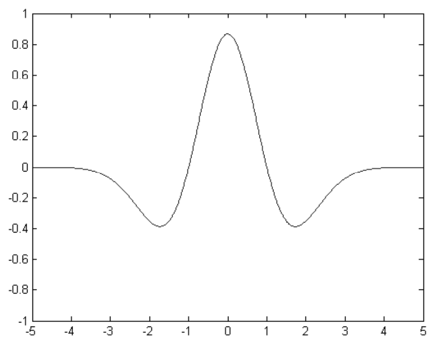

FIGURE 2. The Mexican Hat-wavelet.

We have already seen that the wavelet can be scaled by a and translated by b, so that it is possible to write ψ as

$$\psi_{b:a} = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right). \tag{7.6}$$

Definition 7.3. ***Basic wavelet***
Let $\psi \in L^2(\mathbb{R})$. If

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\psi(\omega)|^2}{|\omega|} d\omega < \infty. \tag{7.7}$$

then ψ is called a basic wavelet.

Definition 7.4. ***Integral Wavelet Transform***
Let $\psi \in L^2(\mathbb{R})$ be a basic wavelet, then the integral wavelet transform W_ψ of a function f is

$$(W_\psi f)(b,a) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t)\overline{\psi\left(\frac{t-b}{a}\right)} dt = \langle f, \psi_{b:a}\rangle \tag{7.8}$$

As with the windowed Fourier transform it is possible to reconstruct the function f if the Integral wavelet transform is known. In order to do so the wavelet has to satisfy (7.7).

Theorem 7.5. *Let ψ be a basic wavelet which defines an integral wavelet transform W_ψ.*
Then for all functions $f, g \in L^2(\mathbb{R})$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[(W_\psi f)(b,a)\overline{(W_\psi g)(b,a)}\right] \frac{da}{a^2} db = C_\psi \langle f, g\rangle.$$

Let, in addition, $f \in L^2(\mathbb{R})$ be continuous in $x \in \mathbb{R}$. Then

$$f(x) = \frac{1}{C_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[(W_\psi f)(b,a)\right] \psi_{b:a}(x)\frac{da}{a^2} db.$$

Proof. First

$$\frac{1}{2\pi}\hat{\psi}_{b:a}(\omega) = \frac{1}{2\pi\sqrt{|a|}}\int e^{-i\omega t}\psi\left(\frac{t-b}{a}\right) dt = \frac{a}{2\pi\sqrt{|a|}} e^{-ib\omega}\hat{\psi}(a\omega).$$

Set

$$F(x) = \hat{f}(x)\hat{\psi}(ax)$$
$$G(x) = \hat{g}(x)\hat{\psi}(ax)$$

and use Theorem 5.18 to derive

$$\begin{aligned}
&\int \left[(W_\psi f)(b,a)\overline{(W_\psi g)(b,a)}\right] db \\
&= \tfrac{1}{|a|}\int \left\{\int f(t)\overline{\psi\left(\tfrac{t-b}{a}\right)}dt \int \overline{g}(s)\psi\left(\tfrac{s-b}{a}\right) ds\right\} db \\
&= \tfrac{a^2}{|a|}\int \left\{\tfrac{1}{2\pi}\overline{\int \bar{F}(x)e^{-ibx}dx}\right\}\left\{\tfrac{1}{2\pi}\int \bar{G}(y)e^{-iby}dy\right\} db \\
&= \tfrac{a^2}{2\pi|a|}\int \bar{G}(x)F(x)dx.
\end{aligned}$$

Use $F(x), G(x)$ as above and C_ψ from (7.7). Then

$$\begin{aligned}
&\int \left\{ \int \left[(W_\psi f)(b,a)\overline{(W_\psi g)(b,a)} \right] db \right\} \tfrac{da}{a^2} \\
&= \tfrac{1}{2\pi} \int \left\{ \hat{f}(x)\overline{\hat{g}(x)} \int \tfrac{|\hat{\psi}(ax)|^2}{|a|} da \right\} dx \\
&= \tfrac{1}{2\pi} \int \left\{ \hat{f}(x)\overline{\hat{g}(x)} \int \tfrac{|\hat{\psi}(y)|^2}{|y|} dy \right\} dx \\
&= \tfrac{1}{2\pi} \left\langle \hat{f}, \hat{g} \right\rangle C_\psi = C_\psi \left\langle f, g \right\rangle .
\end{aligned}$$

When f is continuous at x it is possible to use $g = g_\alpha$ and letting $\alpha \to 0^+$ in order to therefore obtain

$$\begin{aligned}
f(x) &= \frac{1}{C_\psi} \lim_{\alpha \to 0^+} \int \int \left[(W_\psi f)(b,a) \overline{\left\langle g_\alpha(\dot{-}x), \psi_{b:a} \right\rangle} \right] \frac{da}{a^2} db \\
&= \frac{1}{C_\psi} \int \int \left[(W_\psi f)(b,a) \right] \psi_{b:a}(x) \frac{da}{a^2} bd.
\end{aligned}$$

□

7.1. **Wavelet series.** For example when working with image processing it is convenient to use discrete samples of the integral wavelet transform. This can be done with wavelet series. Before the wavelet series can be defined we need different classes of wavelets.

Definition 7.6. ***Orthonormal and Semi-orthonormal wavelets*** *Let $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$ be an R-function as in Definition 3.12 that generates a basis $\{\psi_{j,k}\}$. Then*

(1) *ψ is an orthonormal wavelet if $\{\psi_{j,k}\}$ satisfies*

$$\langle \psi_{j,k}, \psi_{l.m} \rangle = \delta_{j,l}\delta_{k,m} \tag{7.9}$$

(2) *ψ is a semi-orthonormal wavelet if $\{\psi_{j,k}\}$ satisfies*

$$\langle \psi_{j,k}, \psi_{l.m} \rangle = 0, \tag{7.10}$$

An orthonormal wavelet is self dual so that $\psi^{j,k} = \psi_{j,k}$.

For the Riesz basis $\{\psi_{j.k}\}$ in Definition 3.10 the dual is denoted by $\left\{\psi^{j.k}\right\}$. We will need the dual of a semi-orthonormal wavelet to be able to define both wavelet series and the important R-wavelet. The dual is given by the following theorem.

Theorem 7.7. *Let ψ be a semi-orthonormal wavelet, and define the dual $\tilde{\psi}$ of ψ by its Fourier transform;*

$$\hat{\tilde{\psi}}(\omega) = \frac{\hat{\psi}(\omega)}{\sum \left| \hat{\psi}(\omega + 2\pi k) \right|^2}. \tag{7.11}$$

Then

$$\left\langle \psi_{j,k}, \tilde{\psi}_{l.m} \right\rangle = \delta_{j,l}\delta_{k,m} \tag{7.12}$$

where

$$\tilde{\psi}_{l,m}(x) = 2^{l/2}\tilde{\psi}(2^l x - m). \tag{7.13}$$

Proof. Note that $\{\psi_{j,k}\}$ is a Riesz basis of $L^2(\mathbb{R})$ since ψ is a semi-orthogonal wavelet.
Let the Riesz bounds be A, B
Let $\{c_{j,k}\}$ be in $\ell^2(\mathbb{Z}^2)$ and let $c_{j,k} = c_k\delta_{j,0}$. Then (3.4) holds with $\beta = \psi$.
Since

$$A \leq \sum |\hat{\psi}(x + 2\pi k)|^2 \leq B$$

the denominator in (7.11) will be bounded which means that $\tilde{\psi}$ is in $L^2(\mathbb{R})$ and if

$$\tilde{\psi}(x) = \sum a_k \psi(x - k)$$

with

$$a_k = \frac{1}{2\pi}\int_0^{2\pi} a^{-ikx}\frac{1}{\sum |\hat{\psi}(x + 2\pi j)|^2}dx$$

then $\{a_k\}$ will be in ℓ^2. By definition of semi-orthogonal wavelets and (7.12) we get

$$\left\langle \psi_{j,k}, \tilde{\psi}_{l,m} \right\rangle = 0 \text{ when } j \neq l.$$

For $j = l$, let $p = k - m$. Then

$$\begin{aligned}
\left\langle \psi_{j,k}, \tilde{\psi}_{j,m} \right\rangle &= 2^j \int \psi(2^j x - k)\overline{\tilde{\psi}(2^j x - m)}dx \\
&= \int \psi(y - p)\overline{\tilde{\psi}(y)}dy \\
&= \tfrac{1}{2\pi} \int e^{-ip\omega}\hat{\psi}(\omega)\overline{\hat{\tilde{\psi}}(\omega)}d\omega \\
&= \tfrac{1}{2\pi} \int_0^{2\pi} e^{-ip\omega} \left(\sum_k \hat{\psi}(\omega + 2\pi k)\overline{\hat{\tilde{\psi}}(\omega)} \right) d\omega \\
&= \tfrac{1}{2\pi} \int_0^{2\pi} e^{-ip\omega} d\omega = \delta_{p,0} = \delta_{k,m}.
\end{aligned}$$

□

In this way the dual basis $\{\psi^{j.k}\}$ can be written as $\left\{\tilde{\psi}_{j.k}\right\}$

By the above theorem it is possible to construct an orthonormal wavelet from a semi-orthonormal wavelet. Simply let

$$\hat{\psi}^{\perp}(\omega) = \frac{\hat{\psi}(\omega)}{\left(\sum \left|\hat{\psi}(\omega + 2\pi k)\right|^2\right)^{1/2}}$$

then the dual $\tilde{\psi}^{\perp}$ of $\psi^{\perp}$ is given by

$$\hat{\tilde{\psi}}^{\perp}(\omega) = \frac{\hat{\psi}^{\perp}(\omega)}{\sum \left|\hat{\psi}^{\perp}(\omega + 2\pi k)\right|^2} = \hat{\psi}^{\perp}(\omega)$$

so that $\psi^{\perp}$ is self dual and hence orthonormal.
It is however not clear that every R-function has a dual such that the Riesz basis $\{\psi_{j.k}\}$ is given by $\left\{\tilde{\psi}_{j.k}\right\}$. But when this is the case we can call the function an R-wavelet.

Definition 7.8. ***R-wavelet***
Let ψ be an R-function with a dual $\tilde{\psi}$ such that $\{\psi_{j.k}\}$ and $\left\{\tilde{\psi}_{j.k}\right\}$ satisfies (7.12). Then ψ is called an R-wavelet.

Example 7.9. *Let $\eta \in L^2(\mathbb{R})$ be any orthogonal wavelet.*
Consider for every z satisfying $|z| < 1$ the function

$$\psi(x) = \psi_z(x) = \eta(x) - \bar{z}\sqrt{2}\eta(2x).$$

$\{\psi_{j,k}\}$ is then a Riesz basis of $L^2(\mathbb{R})$.
Let $\left\{\psi^{j,k}\right\}$ be the dual basis of $\{\psi_{j,k}\}$ and consider

$$\psi^{0,0}(x) = \sum_{j=0}^{\infty} \eta_{-j,0}(x) z^j \text{ and } \psi^{0,1}(x) = \eta_{0,1}(x).$$

Assume that there exists a function $\tilde{\psi} = \tilde{\psi}_z \in L^2(\mathbb{R})$ such that

$$\psi_z^{j,k} = \tilde{\psi}_z \text{ and } \tilde{\psi}_{j,k} = 2^{j/2}\tilde{\psi}(2^j x - k).$$

Then

$$\begin{aligned} \eta(x) = \ & \eta_{0,1}(x+1) = \psi^{0,1}(x+1) \\ & = \tilde{\psi}_{0,1}(x+1) = \tilde{\psi}_{0,0}(x) = \psi^{0,0}(x) = \textstyle\sum_{j=0}^{\infty} \eta_{-j,0}(x) z^j \end{aligned}$$

and

$$\sum_{j=0}^{\infty} \eta_{-j,0}(x) z^j = 0.$$

But this only holds for some values of z in $|z| < r$, $0 < r < 1$, so $\tilde{\psi}$ does not exist in general.

If $\tilde{\psi}$ does exist, it is an R-wavelet itself.

Now it is time to introduce the wavelet series.

Definition 7.10. ***Wavelet series***
If ψ is an R-function with dual $\tilde{\psi}$. Then

$$f = \sum_{j,k\in\mathbb{Z}} c_{j,k}\psi_{j,k}(x) = \sum_{j,k\in\mathbb{Z}} d_{j,k}\tilde{\psi}_{j,k}(x) \tag{7.14}$$

with

$$c_{j,k} = \left\langle f, \tilde{\psi}_{j,k} \right\rangle \text{ and } d_{j,k} = \langle f, \psi_{j,k} \rangle$$

is the wavelet series of f.

This is a way to discretize the integral wavelet transform which gives more possibilities to recover a function f from its transform. In $L^2(\mathbb{R})$ the functions can actually be recovered from either $\{c_{j.k}\}$ or $\{d_{j.k}\}$ as shown by the following theorem.

Theorem 7.11. *Let ψ be a wavelet with the dual $\tilde{\psi}$. Let $f \in L^2(\mathbb{R})$ and let ψ and $\tilde{\psi}$ be basic wavelets on $(a,b) = \left(\frac{k}{2^j}, \frac{1}{2^j}\right)$. Then the Integral Wavelet transform gives*

$$d_{j,k} = \langle f, \psi_{j,k} \rangle = (W_\psi f)\left(\frac{k}{2^j}, \frac{1}{2^j}\right)$$

$$c_{j,k} = \left\langle f, \tilde{\psi}_{j,k} \right\rangle = (W_{\tilde{\psi}} f)\left(\frac{k}{2^j}, \frac{1}{2^j}\right)$$

and f can be recovered from $d_{j,k}$ or $c_{j,k}$

An informal proof of this statement can be found in [3].

For two functions $f, g \in L^2(\mathbb{R})$ the inner product can be recovered from a discrete sample of the integral wavelet transform as follows

$$\langle f, g \rangle = \sum_{j,k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle \left\langle \tilde{\psi}_{j,k}, g \right\rangle.$$

8. Multiresolution Analysis

Multiresolution analysis is used to construct and decompose wavelets and to decompose signals into wavelet coefficients which makes it useful for example when compressing the data of a photo.
Consider a portrait with a single coloured background. The background does not need the same grade of resolution as the areas around the person and with multiresolution, or MRA, it is possible to divide the photo into different areas and use different resolution for those areas so that the data can be compressed.

Definition 8.1. ***MRA***
A multiresolution analysis (MRA) is defined as a sequence of nested spaces V_j satisfying

1. $\{0\} \leftarrow ...V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset ... \rightarrow L^2(\mathbb{R})$
2. $\overline{\bigcup_{j \in \mathbb{Z}} V_j} = L^2(\mathbb{R})$
3. $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$
4. $f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1}$
5. $V_{j+1} = V_j + W_j$ *when* $j \in \mathbb{Z}$
6. $f(x) \in V_j \Leftrightarrow f(x + \frac{1}{2^j}) \in V_j$.

Definition 8.2. ***Scaling function***
Let $\phi_{j,k}(x) = 2^{j/2}(2^j x - k)$ be in $L^2(\mathbb{R})$ such that it satisfies the properties 1,4,5 and 6 above. Let ϕ generate a Riesz basis $\{\phi_{0,k} : k \in \mathbb{Z}\}$ of V_0. Then ϕ is called a scaling function.

In an MRA there is a scaling function ϕ associated to every V_j.
In this section we will first use a scaling function to generate an MRA and then show how to construct a wavelet basis from the MRA.
Assume that there exists a scaling function ϕ defined as in 8.2. If

$\phi_{j,k}(x)$ generate a closed subspace V_j then this subspace satisfies the properties 1, 4, 5 and 6 in definition 8.1. The following propositions show that they also satisfies properties 2 and 3 and hence generate an MRA.

Proposition 8.3. *Let ϕ be such that*

$$\phi(x) = \sum_k c_k \phi(2x - k) \tag{8.1}$$

with $\sum_k |c_k|^2 < \infty$ and

$$0 < A \leq \sum_{j \in \mathbb{Z}} \left| \hat{\phi}(\omega + 2\pi j) \right|^2 \leq B < \infty. \tag{8.2}$$

Set $V_j = \overline{span\,\{\phi_{j,k} : k \in \mathbb{Z}\}}$. Then

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$$

Proof. By (8.2) we can see that $\phi_{0,k}$ give a Riesz basis of V_0, which also means that they constitute a frame of V_0. By definition of frames then exist A and B such that

$$A \|f\|^2 \leq \sum_{k \in \mathbb{Z}} | \langle f, \psi_{0,k} \rangle |^2 \leq B \|f\|^2 \tag{8.3}$$

Because of the scaling property of the spaces V_j, there exists a mapping D_j sending V_0 onto V_j so that

$$A \|f\|^2 \leq \sum_{k \in \mathbb{Z}} | \langle f, \psi_{j,k} \rangle |^2 \leq B \|f\|^2 \tag{8.4}$$

holds for all f in V_j.
Take an $f \in \bigcap_{j \in \mathbb{Z}} V_j$ and choose an ϵ arbitrarily small. Then there is a continuous $\tilde{f}$ such that

$$\left\| f - \tilde{f} \right\|_{L\infty} \leq \epsilon. \tag{8.5}$$

Let P_j be the orthogonal projection on V_j. Since

$$\left\| f - P_j \tilde{f} \right\| = \left\| P_j (f - \tilde{f}) \right\| \leq \left\| f - \tilde{f} \right\| \leq \epsilon$$

we have

$$\|f\| \leq \epsilon + \left\| P_j \tilde{f} \right\| \text{ for all } j \in \mathbb{Z}. \tag{8.6}$$

We can conclude from the properties of frames that
$\left\| P_j \tilde{f} \right\| \leq \frac{1}{\sqrt{A}} \left(\sum_{k \in \mathbb{Z}} | \left\langle \tilde{f}, \phi_{j,k} \right\rangle |^2 \right)^{1/2}$.

Set $S_{R,j} = \bigcup_{k\in\mathbb{Z}}[k - \frac{R}{2^j}, k + \frac{R}{2^j}]$ with j so large that $\frac{R}{2^j} \leq \frac{1}{2}$. Then

$$\begin{aligned}&\sum |\left\langle \tilde{f}, \phi_{j,k}\right\rangle| \\ &\leq \tfrac{1}{2^j}\sum_k \left(\int_{|x|\leq R}|\tilde{f}(x)||\phi(2^jx-k)|dx\right)^2 \\ &\leq \tfrac{1}{2^j}\left\|\tilde{f}\right\|^2_{L^\infty}\sum_k\int_{|x|\leq R}|\phi(2^jx-k)|^2dx \\ &= \left\|\tilde{f}\right\|^2_{L^\infty} 2R\int_{S_{R,j}}|\phi(y)|^2dy.\end{aligned} \tag{8.7}$$

Let $\chi_j(y)$ be a function such that $\chi_j(y) = 1$ when $y \in S_{R,j}$ and 0 otherwise, then the last part of equation (8.7) can be written as

$$\sum |\left\langle \tilde{f}, \phi_{j,k}\right\rangle| \leq 2R\left\|\tilde{f}\right\|^2\int_{\mathbb{R}}\chi_j(y)|\phi(y)|^2dy. \tag{8.8}$$

For all non-integer j we have $\chi_j \to 0$ as $j \to \infty$, so (8.8) tends to 0 as $j \to \infty$.
There is a j such that (8.7)$\leq \epsilon^2 A$. Combining this with (8.6) gives that $\|f\| \leq 2\epsilon$ and since ϵ was chosen arbitrarily small this shows that $f = 0$ so that $\bigcap_{j\in\mathbb{Z}} V_j = 0$. □

Proposition 8.4. *Take $\phi \in L^2(\mathbb{R})$ satisfying (8.2).*
Let $\hat{\phi}(x)$ be bounded for all x and continuous near $x = 0$ with $\hat{\phi}(0) \neq 0$.
Let V_j be as in Proposition 8.3. Then

$$\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R}).$$

Proof. As in the previous proof $\phi_{j,k}$ build a frame and a Riesz basis with the bounds A and B independent of j.
Take an $f \in \left(\bigcup_{j\in\mathbb{Z}} V_j\right)^{\perp}$ and fix an ϵ arbitrarily small.
For this ϵ there exists an compactly supported function $\tilde{f} \in C^\infty$ such that $\left\|f - \tilde{f}\right\|_{L^2} \leq \epsilon$, so for all $J = -j \in \mathbb{Z}$

$$\left\|P_{-J}\tilde{f}\right\| = \left\|P_j\tilde{f}\right\| = \left\|P_j(\tilde{f} - f)\right\| \leq \epsilon$$

Then

$$\left\|P_{-J}\tilde{f}\right\|^2 \geq B^{-1}\left|\left\langle \tilde{f}, \phi_{-J,k}\right\rangle\right|^2. \tag{8.9}$$

Let $|R| \leq \left\|\hat{\phi}\right\|^2_{L^\infty}\sum_{j\neq 0}\int|\hat{\tilde{f}}(x)||\hat{\tilde{f}}(x + 2^J\pi j)|$, then

$$\sum\left|\left\langle \tilde{f}, \phi_{-J,k}\right\rangle\right|^2 = 2\pi\int|\hat{\phi}(2^{-J}x)|^2|\hat{\tilde{f}}(x)|^2 + R. \tag{8.10}$$

Since $f \in C^\infty$ it is possible to find a konstant C so that

$$|\tilde{f}(x)| \leq C(1 + |x|^2)^{-3/2} \tag{8.11}$$

and when using

$$\sup_{x,y\in\mathbb{R}} (1+y^2)[1+(x-y)^2]^{-1}[1+(x+y)^2]^{-1} < \infty$$

this gives

$$\begin{aligned} |R| &\le C^2 \left\|\hat{\phi}\right\|_{L^\infty}^2 \sum_{j\neq 0} \int (1+|x+2^J\pi j|^2)^{-3/2}(1+|x-2^J\pi j|^2)^{-3/2}dx \\ &\le C' \left\|\hat{\phi}\right\|_{L^\infty}^2 \sum_{j\neq 0} (1+\pi^2 j^2 2^{2J})^{-1/2} \int (1+|y|^2)^{-1}dy \\ &\le C'' 2^{-J}. \end{aligned} \tag{8.12}$$

Combining the equations (8.9),(8.10),(8.11) and (8.12) we get

$$2\pi \int |\hat{\phi}(2^{-J}x)|^2 |\hat{\tilde{f}}(x)|^2 \le B\epsilon^2 + C''2^{-J}. \tag{8.13}$$

The function $\hat{\phi}(x)$ is uniformly bounded and continuous at $x = 0$, so the left-hand side of equation(8.13) tends to $2\pi|\hat{\phi}(0)|^2 \left\|\tilde{f}\right\|_{L^2}^2$ as $J \to \infty$. For a C independent of ϵ we have $\left\|\tilde{f}\right\|_{L^2} \le |\hat{\phi}(0)|^{-1}C_\epsilon$ so that

$$\|f\|_{L^2} \le \epsilon + \left\|\tilde{f}\right\|_{L^2} \le (1 + C|\hat{\phi}(0)|^{-1})\epsilon.$$

ϵ was chosen arbitrarily small so $f = 0$, and since $f \in \left(\bigcup_{j\in\mathbb{Z}} V_j\right)^\perp$ this means that $\left(\bigcup_{j\in\mathbb{Z}} V_j\right)^\perp \to 0$ as $j \to \infty$ which is equivalent to $\left(\bigcup_{j\in\mathbb{Z}} V_j\right) \to L^2(\mathbb{R})$ as $j \to \infty$. $\square$

As mentioned above this shows that a scaling function ϕ generates a multiresolution analysis V_j of $L^2(\mathbb{R})$. Next we will show that if there exists an MRA, then there exists an orthonormal wavelet basis of $L^2(\mathbb{R})$.

If ϕ generates a Riesz basis of V_0 then property 6 in definition 8.1 ensures that $\{\phi_{j,k} : k \in \mathbb{Z}\}$ is a Riesz basis of any V_j. Recall Example 2.27 with the nested subspaces V_j and the mutually orthogonal subspaces W_j. Whenever there exists a sequence of such subspaces as V_j it is possible to find an orthonormal wavelet basis $\{\psi_{j,k} : j,k \in \mathbb{Z}\}$ of $L^2(\mathbb{R})$ generated by

$$\psi_{j,k} = 2^{j/2}\psi(2^j x - k). \tag{8.14}$$

This generates mutually orthonormal subspaces W_j such that a direct sum decomposition of $L^2(\mathbb{R})$ can be written as

$$L^2(\mathbb{R}) = \oplus_{j\in\mathbb{Z}} W_j. \tag{8.15}$$

Also the properties that

$$\phi(x) \in V_j \subset V_{j+1}$$

and

$$\psi(x) \in W_j \subset V_{j+1}$$

makes it possible to express ϕ and ψ in the basis of V_{j+1}. Hence there exist sequences $\{p_k\}$ and $\{q_k\}$ such that the scaling function and the wavelet can be written

$$\phi(x) = \sum_{k=-\infty}^{\infty} p_k \phi(2k - x)$$

and

$$\psi(x) = \sum_{k=-\infty}^{\infty} q_k \phi(2k - x).$$

These sequences are called the **two-scale relations**.

Example 8.5. *If ϕ is the Haar scaling function*

$$\phi(x) = \begin{cases} 1 & \text{if } 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{8.16}$$

then the Haar wavelet is given by

$$\frac{1}{\sqrt{2}}\psi(\frac{x}{2}) = \sum (-1)^{1-k} p_{(1-k)} \phi(x-k) = \frac{1}{\sqrt{2}} (\phi(x-1) - \phi(x))$$

where the two-scale sequence p_k is given by

$$p_k = \begin{cases} \frac{1}{\sqrt{2}} \text{ when } k = 0, 1 \\ 0 \text{ otherwise } . \end{cases}$$

The relation can also be expressed as

$$\psi(x) = \phi(x) - \phi(2x - k),$$

giving the function ψ in (2.15)

Example 8.6. *For the class of m^{th} order cardinal B-splines defined in Example 6.10, the two-scale relation is given by*

$$N_m(x) = \sum_{k=0}^{\infty} 2^{-m+1} \binom{m}{n} N_m(2x - k),$$

with

$$p_k = 2 \int_{-\infty}^{\infty} \phi(x) \overline{\phi(2x - k)} dx.$$

These assumptions are made regarding the scaling function ϕ and the corresponding $\{p_k\}$;

- $\phi \in L^1(\mathbb{R})$
- $\sum_{k=-\infty}^{\infty} \phi(x - k) = 1$
- $\{p_k\} \in \ell^1$
- $\sum p_{2k} = \sum p_{2k+1} = 1$

Next the two-scale symbol will be introduced to show some simplifications that can be made in the construction of the wavelet spaces.

Definition 8.7. ***Two-scale symbol***
Let $\{p_k\}$ define a two-scale relation for ϕ. Then

$$P(z) = \frac{1}{2} \sum_{k=-\infty}^{\infty} p_k z^k \tag{8.17}$$

is called the two-scale symbol of ϕ.

The Fourier transform of ϕ can be written by means of (8.17) as

$$\hat{\phi}(\omega) = P(z)\hat{\phi}(\frac{\omega}{2}).$$

Example 8.8. *For the Haar scaling function in (2.14) we get $p_0 = p_1 = 1$ and $p_k = 0$ for $k \neq 0, 1$ so that the two-scale symbol becomes*

$$P(z) = \frac{(1+z)}{2}.$$

This gives the Fourier transform

$$\hat{\phi}(\omega) = P(e^{-\frac{i\omega}{2}})\hat{\phi}(\frac{\omega}{2}) = \frac{1}{2}(1 + e^{-i\omega/2})\left(\frac{e^{-i\omega/2} - 1}{-i\sqrt{2\pi}\omega/2}\right) = \frac{e^{-i\omega} - 1}{-\sqrt{2\pi}i\omega}.$$

When the sequence $\{p_k\}$ is in ℓ^1 and the two scale symbol $P_\phi(z)$ is given by (8.17) we can associate another sequence $\{q_k\} \in \ell^1$ with the corresponding symbol

$$Q(z) = \frac{1}{2} \sum_{k=-\infty}^{\infty} q_k z^k.$$

as follows. If ϕ generates the space V_j as described above, then ψ generates the space W_j. Thus when P_ϕ gives the relation $V_j \subset V_{j+1}$ and Q gives $W_j \subset V_{j+1}$ this ensures that

$$V_j \cap W_j = 0 \text{ and } V_{j+1} = V_j \oplus W_j.$$

Example 8.9. *The function ψ can also be written (with $q_k = (-1)^k \overline{p_{1-k}}$)*

$$\psi(x) = \sum q_k = (-1)^k \overline{p_{1-k}} \phi(2x - k).$$

For the Haar case one has $p_0 = p_1 = 1$ which gives

$$\psi(x) = \phi(x) - \phi(2x - k)$$

as in Example 8.5.

From property 5 in Definition 8.1 it is clear that W_j is the orthogonal complement to V_j in V_{j+1} (see [4], page 131) W_j also have the scaling property 4. Hence in order to find a basis for any W_j it is only necessary to find a function ψ that generates a basis for W_0 .
In other words, if $\{\psi_{0,k} :\in \mathbb{Z}\}$ is an orthonormal basis for W_0 then $\{\psi_{j,k} :\in \mathbb{Z}\}$ will be an orthonormal basis for W_j.

Before we can continue and show when a wavelet is a basis for $L^2(\mathbb{R})$ we have to define filters. Recall the notion of the continuous convolution in definition 5.15. The discrete version as follows;

Definition 8.10. ***Discrete Convolution***
Let $h_k = \frac{1}{2}(-1)^k p_{k+1}$ and $g_k = \frac{1}{2}\overline{p_k}$ be two absolutely convergent sequences, then the discrete convolution $Lg = h * g$ is defined as

$$(h * g)_k = \sum_{n\in\mathbb{Z}} h_{k-n} g_n. \tag{8.18}$$

A filter is defined as a convolution operator. An example is the discrete convolution Lg above which is a discrete filter.

Definition 8.11. ***Impulse response function***
Let $Lf = f * h$ be a filter. Then h(x) is called the impulse response function.

Example 8.12. *Let $Lf(x) = \int_{-\infty}^{infy} h(t)f(x-t)$ be a continuous convolution as in definition (5.15). If $Lf(x)$ does not depend on the values $f(x)$ for $x > t$ then $h(x) = 0$ for $x < 0$ and $Lf(x)$ is said to be a causal filter. A causal filter is defined as a filter where the output signal begins after the input signal has started to arrive.*

Example 8.13. *In signal processing a low-pass filter is used to filter out frequencies above a certain "'cut-off frequency"'. An ideal low-pass filter leaves the frequencies below this cut-off unchanged but gets rid all frequencies above. The impulse response function of the ideal low-pass filter is given by*

$$h(x) = \frac{1}{2\pi}\int_{-a}^{a} e^{i\omega x} d\omega = \frac{sin(ax)}{\pi x}.$$

The filter is a convolution with a sine function in the time domain or a multiplication with a rectangular function (like the Haar wavelet) in the frequency domain.

Definition 8.14. ***Conjugate mirror filter***
Let Lf be a discrete filter and let the impulse response function $h(x)$ satisfy

$$|\hat{h}(\omega)|^2 + |\hat{h}(\omega+\pi)|^2 = 2.$$

Then Lf is called a conjugate mirror filter.

Theorem 8.15. *Let ϕ be a scaling function and let h be its conjugate mirror filter.*
Let ψ be defined by its Fourier transform as

$$\hat{\psi}(\omega) = \frac{1}{\sqrt{2}}\hat{g}(\frac{\omega}{2})\hat{\phi}(\frac{\omega}{2}) \text{ with } \hat{g}(\omega) = e^{-i\omega}\overline{\hat{h}(\omega+\pi)}. \tag{8.19}$$

If $\psi_{j,k}(x) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t-2^j n}{2^j}\right)$ for any scale 2^j then $\{\psi_{j,k} : k \in \mathbb{Z}\}$ is an orthonormal basis for W_j and $\{\psi_{j,k} : j,k \in \mathbb{Z}\}$ is an orthonormal basis of $L^2(\mathbb{R})$ for all scales.

Proof. First, it is possible to write $\hat{\psi}(\omega)$ as in (8.19) since: $\psi(\frac{t}{2}) \in W_1 \subset V_0$ so that the decomposition $\{\phi(t-n)\}_{n\in\mathbb{Z}}$ is an orthonormal basis of V_0. Let

$$\frac{1}{\sqrt{2}}\psi(\frac{t}{2}) = \sum_{-\infty}^{\infty} g_n\phi(t-n) \text{ where } g_n = \frac{1}{\sqrt{2}}\left\langle \psi\left(\frac{t}{2}\right), \phi(t-n)\right\rangle \tag{8.20}$$

then the Fourier transform of it becomes

$$\hat{\psi}(2\omega) = \frac{1}{\sqrt{2}}\hat{g}(\omega)\hat{\phi}(\omega). \tag{8.21}$$

To prove the rest we start with the following statement:

Lemma 8.16. *The family* $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ *is an orthonormal basis of* W_j *if and only if*

$$|\hat{g}(\omega)|^2 + |\hat{g}(\omega+\pi)|^2 = 2 \tag{8.22}$$

and

$$\hat{g}(\omega)\overline{\hat{h}(\omega)} + \hat{g}(\omega+\pi)\overline{\hat{h}(\omega+\pi)} = 0. \tag{8.23}$$

Proof. This proof will be made for $j = 0$, then it can be extended by scaling to hold for any j. The sequence $\{\psi(t-n)\}_{n\in\mathbb{Z}}$ is orthonormal if and only if, for all $\omega \in \mathbb{R}$

$$\sum_n |\hat{\psi}(\omega + 2\pi n)|^2 = 1. \tag{8.24}$$

Since $\hat{g}(\omega)$ is 2π-periodic and $\hat{\psi}(\omega) = \frac{1}{\sqrt{2}}\hat{g}(\frac{\omega}{2})\hat{\phi}(\frac{\omega}{2})$:

$$\begin{aligned}&\textstyle\sum_n |\hat{\psi}(\omega+2\pi n)|^2 = \sum_n^{\infty} |\hat{g}\left(\frac{\omega}{2}+\pi n\right)|^2 |\hat{\phi}\left(\frac{\omega}{2}+\pi n\right)|^2 \\ &= |\hat{g}\left(\tfrac{\omega}{2}\right)|^2 \textstyle\sum_p^{\infty} |\hat{\phi}\left(\frac{\omega}{2}+2\pi p\right)|^2 + |\hat{g}\left(\frac{\omega}{2}+\pi\right)|^2 \sum_p^{\infty} |\hat{\phi}\left(\frac{\omega}{2}+\pi+2\pi p\right)|^2.\end{aligned}$$

Since $\sum |\hat{\phi}(\omega+2\pi p)|^2 = 1$, (8.24) satisfies (8.22).

V_0 and W_0 are orthogonal if

$$\langle \psi(t), \phi(t-n)\rangle = \psi * \overline{\phi(n)} = 0$$

The Fourier transform, $\hat{\psi}\overline{\hat{\phi}(t)}$ is equal to 0 if

$$\sum \hat{\psi}(\omega+2\pi n)\overline{\hat{\phi}(\omega+2\pi n)} = 0 \tag{8.25}$$

Let

$$\hat{\psi}(\omega) = \frac{1}{\sqrt{2}}\hat{g}(\frac{\omega}{2})\hat{\phi}(\frac{\omega}{2})$$

and

$$\hat{\phi}(\omega) = \frac{1}{\sqrt{2}}\hat{h}(\frac{\omega}{2})\hat{\phi}(\frac{\omega}{2})$$

in (8.25), then there is equality between (8.23) and (8.25). The sequence $\left\{\sqrt{2}\phi(2t-n)\right\}_{n\in\mathbb{Z}}$ is an orthonormal basis of V_1.

To show that $V_1 = V_0 \oplus W_0$ is the same as showing that for any $a_n \in \ell^2(\mathbb{Z})$ there exists a $b_n \in \ell^2(\mathbb{Z})$ and a $c_n \in \ell^2(\mathbb{Z})$ so that

$$\sum_n a_n \sqrt{2}\phi(2(t-\frac{n}{2})) = \sum_n b_n \phi(\pi - n) + \sum_n c_n \psi(t-n). \tag{8.26}$$

The Fourier transform of the above becomes

$$\frac{1}{\sqrt{2}}\hat{a}\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right) = \hat{b}(\omega)\hat{\phi}(\omega) + \hat{c}(\omega)\hat{\psi}(\omega)$$

To be able to insert the same $\hat{\psi}(\omega)$ and $\hat{\phi}(\omega)$ as before in this Fourier transform it is necessary that the following relation holds:

$$\hat{a}\left(\frac{\omega}{2}\right) = \hat{b}(\omega)\hat{h}\left(\frac{\omega}{2}\right) + \hat{c}(\omega)\hat{g}\left(\frac{\omega}{2}\right). \tag{8.27}$$

(8.27) holds when inserting (8.22) and (8.23) together with

$$|\hat{h}(\omega)|^2 + |\hat{h}(\omega+\pi)|^2 = 2. \tag{8.28}$$

By letting

$$\hat{b}(2\omega) = \frac{1}{2}\left(\hat{a}(\omega)\overline{\hat{h}(\omega)} + \hat{a}(\omega+\pi)\overline{\hat{h}(\omega+\pi)}\right)$$

and

$$\hat{c}(2\omega) = \frac{1}{2}\left(\hat{a}(\omega)\overline{\hat{g}(\omega)} + \hat{a}(\omega+\pi)\overline{\hat{g}(\omega+\pi)}\right)$$

we have two 2π-periodic sequences that are the Fourier series of b_n and c_n in (8.26). Thus it is proved that such sequences exist. $\square$

Because of (8.28) $\hat{g}(\omega) = e^{-i\omega}\overline{\hat{h}(\omega+\pi)}$ satisfies (8.22) and (8.23). So Lemma 8.16 says that the function $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ is an orthogonal basis of W_j.
Here

$$V_J = \oplus_{j=-\infty}^{J-1} W_j.$$

And V_j generates a multiresolution analysis, so

$$V_J \to \begin{cases} \{0\} \text{ as } J \to -\infty \\ L^2(\mathbb{R}) \text{ as } J \to \infty. \end{cases}$$

The $W_j's$ are orthonormal and $\oplus_{-\infty}^{\infty} W_j = L^2(\mathbb{R})$, so a union of orthonormal bases of all W_j's is an orthonormal basis of $L^2(\mathbb{R})$. $\square$

This theorem shows that if there exists an MRA, then there exists an orthonormal wavelet basis of $L^2(\mathbb{R})$.

8.1. **Splines and MRA.** A special kind of scaling functions which are very useful when constructing wavelets are called splines.
Splines are piecewise polynomials which match together smoothly. How smoothly they match depends on the order of the spline.
Let us say that we want to approximate a function x on [a,b] by a function y. If we use splines we can represent y by polynomials in every subinterval of a partition of [a,b] where we have one polynomial for every subinterval. So instead of approximating x by one single polynomial, we approximate it by as many as we have intervals. In this way some analyticity may be lost, but instead a good approximation is gained.
The breakpoints between the polynomials are called nodes and a spline of order m consists of polynomials of order m between the nodes.
A special kind of splines that will be considered here are called Cardinal splines, they are splines with equal spacing between the nodes. This section aims to define cardinal B-splines, which for example will be used for the Daubechies wavelets in section 10.1.

Definition 8.17. ***The space*** π_n
Let π_n be the space of all polynomials of degree at most n.

Definition 8.18. ***The space*** S_m
For all $m > 0$
Define S_m as the space of cardinal splines of order m and with the node-sequence $\mathbb{Z}$.
In other words S_m is the collection of all $f \in C^{m-2}$ such that the restriction of f to any interval $[k, k+1)\,, k \in \mathbb{Z}$ is in the space π_{m-1}.

Example 8.19. *A basis for S_1 is $\{N_1(x-k) : k \in \mathbb{Z}\}$, where*

$$N_1(x) = \begin{cases} 1 & x \in [0,1) \\ 0 & \textit{otherwise.} \end{cases} \tag{8.29}$$

If $T = \left\{(x-k)_+^{m-1}\right\}$ is a basis for S_m, then none of the $f \in T$ is in $L^2(\mathbb{R})$. It is however possibly to use functions from T and create $L^2(\mathbb{R})$-functions, by using differences.

$$\begin{cases} (\Delta f)(x) = f(x) - f(x-1) \\ (\Delta^n f)(x) = (\Delta n - 1(\Delta f))(x) \end{cases} \tag{8.30}$$

So that when $f \in \pi_{m-1}$, then $\Delta^m f = 0$.
It is now possible to construct the space of cardinal splines.

Definition 8.20. ***The space*** $\boldsymbol{N}_m$
Let N_1 be as in (8.29).
For $m \geq 2$, let

$$N_m(x) = \frac{1}{(m-1)!}\Delta^m x_+^{m-1} \tag{8.31}$$

where N_m is a linear combination of T.
This can also be written as

$$N_m(x) = \frac{1}{(m-1)!} \sum_{k=0}^{m} (-1)^k \left(\begin{array}{c} m \\ k \end{array} \right) (x-k)_+^{m-1}. \qquad (8.32)$$

N_m has compact support ($\text{supp}N_m = [0, m]$), so it belongs to $L^2(\mathbb{R})$.

Definition 8.21. ***Cardinal B-spline***
Let B be the space $B = \{N_m(x-k) : k \in \mathbb{Z}\}$ of integer translations of N_m. Then B is a basis of S_m and any spline in B is called a cardinal B-spline.

Le us consider the space S_m. For cardinal splines with the node-sequence $2^{-j}\mathbb{Z}$ the corresponding space is called S_m^j, where

$$... \subset S_m^{-1} \subset S_m^0 \subset S_m^1 \subset ... \qquad (8.33)$$

Let $V_j^m = \overline{S_m^j \cap L^2(\mathbb{R})}$, and construct a sequence

$$... \subset V_{-1}^m \subset V_0^m \subset V_1^m \subset ...$$

of closed cardinal spline subspaces.

Then, when B is a basis of V_0^m, the sequence

$$\left\{2^{j/2} M_m(2^j x - k) : k \in \mathbb{Z}\right\}$$

is a Riesz basis for V_j^m with the same Riesz bounds as the basis B.

8.1.1. *Properties of B-splines.* For the proofs of the properties in this section, please see [3].

Definition 8.22. *For $m \geq 2$ the m^{th}-order cardinal B-spline is defined by*

$$N_m(x) = (N_{m-1} * N_1)(x) = \int_0^1 N_{m-1}(x-t)dt. \qquad (8.34)$$

Theorem 8.23. ***Properties of B-splines***

(1) *For $f \in C$*
$$\int_{-\infty}^{\infty} f(x) N_m(x) dx = \int_0^1 ... \int_0^1 f(x_1 + ... + x_m) dx_1 ... dx_m.$$
(2) *For $g \in C^m$*
$$\int_{-\infty}^{\infty} g^{(m)}(x) N_m(x) dx = \sum_{k=0}^{m} (-1)^{m-k} \left(\begin{array}{c} m \\ k \end{array} \right) g(k) \qquad (8.35)$$
(3) *$\text{supp}N_m = [0, m]$, so that $N_m(0) = N_m(m) = 0$*
(4) $N_m'(x) = (\Delta N_{m-1})(x) = N_{m-1}(x) - N_{m-1}(x-k)$
(5) $\sum_k N_m(x-k) = 1$ *for all x.*

(6) *The relation between N_m and N_{m-1} is given by*

$$N_m(x) = \frac{x}{m-1} N_{m-1}(x) + \frac{m-x}{m-1} N_{m-1}(x-1).$$

(7) *N_m is symmetric with respect to the center of its support. That is*

$$N_m\left(\frac{m}{2} - x\right) = N_m\left(\frac{m}{2} + x\right).$$

The properties of the Fourier transform together with the fact that

$$N_m(x) = N_1 * \ldots * N_1(x)$$

and

$$\hat{N}_1(\omega) = \int_0^1 e^{-i\omega t} dt = \frac{1 - e^{-i\omega}}{i\omega},$$

makes it possible to express the Fourier transform of $N_m(\omega)$ as

$$\hat{N}_m(\omega) = \left(\frac{1 - e^{-i\omega}}{i\omega}\right)^m,$$

Example 8.24. *In view of the two-scale symbol in (8.17) we have for the m^{th}-order cardinal B-spline a relation:*

$$P(z) = \left(\frac{1+z}{2}\right)^m$$

and

$$\Pi_{k=1}^n P(e^{-i\omega/2^k}) = \frac{1}{2^{mn}} \left(\frac{1 - e^{-i\omega}}{1 - e^{-i\omega/2^k}}\right)^m,$$

where the right-hand side tends to $\left(\frac{1-e^{-i\omega}}{i\omega}\right)^m$ as $n \to \infty$ which agrees with the above.

When using frames and R-wavelets it is necessary to consider properties such as

- the size of the time-frequency window
- the complexity and efficiency of the function
- the smoothness and symmetry of the function
- the order of approximation needed.

B-splines are a very efficient group of functions since they are the least complicated functions with the smallest support ([2]) so that an MRA constructed from a spline becomes a smooth approximation that has fast decay. (In other words; it disappears fast outside its support.) Consider V_j as the space of splines with degree $m \geq 0$. They are $m-1$ times continuous differentiable functions which behave like a polynomial at degree m inside each interval $[2^j n, 2^j(n+1)]$.
When $m = 0$, a polynomial is piecewise constant and when $m = 1$ the MRA becomes piecewise linear and continuous.

Example 8.25. *Let N_m be the m^{th} order cardinal B-spline. Then the wavelet ψ corresponding to N_m is given by*

$$\psi_m(x) = \sum q_n N_m(2x - n)$$

where q_n is a convolution

$$q_n = \frac{(-1)^n}{2^{m-1}} \sum \binom{m}{l} N_{2m}(n + 1 - l).$$

Example 8.26. *A special class of wavelets are the Battle-Lemarié wavelets. They are constructed via MRA with B-splines as scaling function. The B-splines in this case have their nodes on the integers.*
Let

$$\phi(x) = \begin{cases} 1 - |x| & \text{when } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\phi(x) = \frac{1}{2}\phi(2x + 1) + \phi(2x) + \frac{1}{2}\phi(2x - 1)$$

and the Fourier transform becomes

$$\phi(\omega) = \frac{1}{\sqrt{2\pi}} \left(\frac{\sin(\omega/2)}{\omega/2} \right)^2$$

so that (8.1) and (8.2) are satisfied. In addition $\phi \in L^1(\mathbb{R})$ and $\int \phi(x)dx = 1$ so that V_j generated by ϕ gives an MRA.
Now the Fourier transform of a wavelet can be expressed as

$$\hat{\psi}(\omega) = \sqrt{3}e^{i\omega/2} \sin^2(\omega/4) \left(\frac{1 + 2\sin^2(\omega/4)}{(1 + 2\cos^2(\omega/2)) + (1 + 2\cos^2(\omega/4))} \right)^{1/2} \hat{\phi}(\omega/2).$$

The Fourier coefficients of the denominator can be calculated and ψ can be written as

$$\psi(x) = \frac{\sqrt{3}}{2} \sum (d_{n+1} - 2d_n + d_{n-1})\phi(2x - n).$$

Spline wavelets are continuous but even though they decay exponentially their scaling and wavelet functions have infinite support.

9. Decomposition and Reconstruction

Assume that you have a signal that you want to denoise, i.e. you need to find a way to remove frequencies that are outside a given interval. To do this the signal has to be decomposed into small components that are associated to different frequencies. When this is done it is possible to select which components to keep, in the denoising case by throwing away components associated to frequencies outside the interval. The last step is to put the remaining components back together again, and the importance of the task is to manage to do this in a way such that the result is close to the original signal.

9.1. **The decomposition algorithm.** The theory of multiresolution analysis will be used to break a function into its W_j-parts, which contains information about the oscillatory behavior of the signal.
A signal $f \in V_j$ can be represented both as

$$f = \sum \langle f, \phi_{j,k} \rangle \phi_{j,k} \tag{9.1}$$

and, since $V_j = V_{j-1} \oplus W_{j-1}$, as

$$f = \sum \langle f, \phi_{j-1,k} \rangle \phi_{j-1,k} + \sum \langle f, \psi_{j-1,k} \rangle \psi_{j-1,k} = f_{j-1} + w_{j-1}. \tag{9.2}$$

For $a_k^j = 2^{j/2} \langle f, \phi_{j,k} \rangle$ equation (9.1) becomes

$$f(x) = \sum a_k^j \phi(2^j x - k)$$

and with $b_k^j = 2^{j/2} \langle f, \psi_{j,k} \rangle$ equation (9.2) becomes

$$f(x) = \sum a_k^{j-1} \phi(2^{j-1} x - k) + \sum b_k^{j-1} \psi(2^{j-1} x - k).$$

The number a^j is called the approximation coefficient and b^j is called the detail coefficient. the number b^j corresponds to the frequency of the function (or signal).

To begin with the decomposition a sequence of V_j has to be chosen. It should fit the information in f as good as possible. The choice of V_j depends on the sampling rate and the MRA used. The best approximation of f from V_j is given by the orthogonal projection

$$P_j f = \sum a_k^J \phi(2^j x - k).$$

When a signal is being sampled, the following so-called quadratic rule is sometimes needed to get a good approximation of $P_j f$.

Theorem 9.1. *Let $\{V_j\}_{j \in \mathbb{Z}}$ be an MRA and let ϕ be its compactly supported scaling function. Then, for continuous $f \in L^2(\mathbb{R})$ and for sufficiently large j*

$$a_k^j = \int_{-\infty}^{\infty} f(x) \overline{\phi(2^j x - k)} dx = f\left(\frac{k}{2^j}\right) \int_{-\infty}^{\infty} \overline{\phi(x)} dx. \tag{9.3}$$

Proof. The function ϕ has compact support so it is only nonzero inside an interval $[-M, M]$.
The interval of integration for a_k^j is $\{x : |2^j x - k| \leq M\}$.
Let $t = 2^j x - k$. Then

$$a_k^j = \int_{-M}^{M} f(2^{-j} t + 2^{-j} k) \overline{\phi(t)} dt.$$

As j increases, the number $2^{-j} t + 2^{-j} k$ gets close to $2^{-j} k$ for $t \in [-M, M]$. But since f is uniformly continuous on every finite interval this means that

$$f(2^{-j} t + 2^{-j} k) \approx f(2^{-j} k) \text{ for all } t \in [-M, M].$$

Hence

$$a_k^j \approx f\left(\frac{k}{2^j}\right)\int_{-M}^{M}\overline{\phi(x)}dx.$$

And since ϕ is zero outside $[-M, M]$, this means that

$$\int_{-M}^{M}\overline{\phi(t)}dt = \int_{-\infty}^{\infty}\overline{\phi(t)}dt.$$

Thus formula (9.3) is obtained. □

The approximation of $P_j f$ gets better the bigger j gets, but the calculations gets heavier at the same time, so what j to use is an optimization problem.
$P_j f$ can also be written as

$$P_j f \approx f_j(x) = \int \overline{\phi(x)}dx \sum f\left(\frac{k}{2^j}\right)\phi(2^j x - k).$$

At the first step we have $f \approx f_j \in V_j$. We also know from our previous arguments that $\sum W_j = L^2(\mathbb{R})$. It is possible to start with f_j and then to decompose it so that for

$$f_{j-1} \in V_{j-1} \text{ and } w_{j-1} \in W_{j-1} : f_j = f_{j-1} + w_{j-1}.$$

Depending on the purpose, the level of decomposition differs. Sometimes, like when detecting singularities, it can be enough to downsample one or two steps. For now the stopping level will be 0.

9.2. **Downsampling.** Put

$$h_k = \frac{1}{2}(-1)^k p_{k-1} \tag{9.4}$$

and

$$l_k = \frac{1}{2}\overline{p_{-k}}. \tag{9.5}$$

Then

$$H(x) = h * x \text{ and } L(x) = l * x$$

becomes two discrete filters where L is a low-pass filter and H is a high-pass filter.
With $x = a^j$ we define the downsampling coefficients as

$$a_l^{j-1} = L(a^j)_{2l} \text{ and } b_l^{j-1} = H(a^j)_{2l}. \tag{9.6}$$

In order to make the data-load smaller, the following downsampling can be used.

Definition 9.2. ***Downsampling***
Let $x = (...x_{-2}, x_{-1}, x_0, x_1, x_2...)$. *Then*

$$Dx = (...x_{-2}, x_0, x_2...)$$

is the downsampling of x *and* D *is called the downsampling operator. This can also be written as:*

$$(Dx)_l = x_{2l} \text{ for all } l \in \mathbb{Z}.$$

Hence:

$$a^{j-1} = D(l * a^j) = DLa^j$$
$$b^{j-1} = D(h * a^j) = DHa^j.$$

Since the stopping level of the downsampling was chosen to be 0, this gives a set including the approximation coefficient at the level 0 $\{a_k^0\}$ and the wavelet coefficients $\left\{b_k^{j'}\right\}$ for $j' = 0, ..., j-1$.

The following example was found in [1] as an exercise and its purpose is to illustrate the decomposition.

Example 9.3. *Let* ϕ *and* ψ *be the Haar scaling and wavelet functions in (2.14) and (2.15) respectively.*
Assume that $\phi(2^j x - k)$ *generate* V_j *and* $\psi(2^j x - k)$ *generate* W_j.
Let f *be defined for* $x \in [0, 1)$ *and let*

$$f(x) = \begin{cases} -1 & 0 \leq x < \frac{1}{4} \\ 4 & \frac{1}{4} \leq x < \frac{1}{2} \\ 2 & \frac{1}{2} \leq x < \frac{3}{4} \\ -3 & \frac{3}{4} \leq x < 1. \end{cases}$$

Then f *will be expressed in terms of the basis for* V_2 *and then decomposed into its component parts in* W_1, W_0 *and* V_0.
In V_2 *the function* f *can be written*

$$f(x) = -\phi(4x) + 4\phi(4x-1) + 2\phi(4x-2) - 3\phi(4x-3).$$

Put

$$\begin{array}{l} \phi(x) = \frac{1}{2}(\psi(2x) + \phi(2x)) \\ \phi(4x-1) = \frac{1}{2}(\phi(2x) - \psi(2x)) \\ \phi(4x-2) = \phi(4(x-\frac{1}{2})) = \frac{1}{2}(\psi(2(x-\frac{1}{2})) + \phi(2(x-\frac{1}{2}))) \\ \phi(4x-3) = \phi(4(x-\frac{1}{2})-1) = \frac{1}{2}(\phi(2(x-\frac{1}{2})) - \psi(2(x-\frac{1}{2}))) \end{array} \tag{9.7}$$

into $f(x)$, *then it can be rearranged so that*

$$f(x) = \frac{3}{2}\phi(2x) - \frac{1}{2}\phi(2x-1) - \frac{3}{2}\psi(2x) + \frac{5}{2}\psi(2x-1).$$

Terms in this expression containing ψ *are already in* W_1, *to get the other components the decomposition process has to continue.*

Use

$$\phi(2x) = \frac{1}{2}(\phi(x) + \psi(x))$$
$$\phi(2x-1) = \frac{1}{2}(\phi(x) - \psi(x))$$

and put this into the latest form of $f(x)$ *which then becomes*

$$f(x) = -\phi(x) - \frac{1}{2}(\psi(x) - \frac{3}{2}\psi(2x) + \frac{5}{2}\psi(2x-1)).$$

Now the term with ϕ *is in* V_0*,* $\psi(x)$ *is in* W_0 *and* $\psi(2x)$ *and* $\psi(2x-1)$ *are in* W_1 *so that those terms are the respective spaces components.*

9.3. **The reconstruction algorithm.** When the decomposition is finished we are left with all the wavelet coefficients $b_k^{j'}$ and we have to decide what to do with them. If the signal should be denoised, then the $b_k^{j'}$ corresponding to wrong frequencies has to be thrown away. If the signal should be compressed then it is possible to remove coefficients that are small. Here "small" is a problem of definition depending of how precise the final result should be.

As in the decomposition we again define two discrete filters

$$\tilde{H} = \tilde{h} * x \text{ with } \tilde{h}_k = \overline{p_{1-k}}(-1)^k$$
$$\tilde{L} = \tilde{l} * x \text{ with } \tilde{l}_k = p_k.$$

Now $\tilde{H}$ and $\tilde{L}$ is the reconstruction high- resp. low-pass filters.
The reconstruction formula is written as

$$a_k^j = \sum_{n\in\mathbb{Z}} \tilde{l}_{k-2n} a_n^{j-1} + \sum_{n\in\mathbb{Z}} \tilde{h}_{k-2n} b_n^{j-1}.$$

If the index were $k-n$ instead of $k-2n$ this would have been a convolution. To recover the odd terms that are missing the upsampling operator can be used.

Definition 9.4. ***Upsampling***
Let $x = (...x_{-2}, x_{-1}, x_0, x_1, x_2...)$*. Then*

$$Ux = x = (...x_{-2}, 0, x_{-1}, 0, x_0, 0, x_1, 0, x_2...).$$

is the upsampling of x*, and* U *is called the upsampling operator.*

With this operator it is possible to write

$$a_k^j = \tilde{l} * (Ua^{j-1}) + \tilde{h} * (Ub^{j-1}) = \tilde{L}Ua^{j-1} + \tilde{H}Ub^{j-1}.$$

The following example will illustrate the reconstruction algorithm.

Example 9.5. *In this example a function* $f \in V_3$ *will be reconstructed from the coefficients*

$$a^{[1]} = [\frac{3}{2}, -1] \ , \ b^{[1]} = [-1, -\frac{3}{2}] \ , \ b^{[2]} = [-\frac{3}{2}, -\frac{3}{2}, -\frac{1}{2}, -\frac{1}{2}].$$

To do this we use the relations

$$a^{[j]}(2k) = a_k^{[j-1]} + b_k^{[j-1]}$$
$$a^{[j]}(2k+1) = a_k^{[j-1]} - b_k^{[j-1]}.$$

To get to V_3, which is the third level, two iterations must be done. First use $a^{[1]}$ and $b^{[1]}$ to get

$$a^{[2]} = [\frac{1}{2}, \frac{5}{2}, -\frac{5}{2}, \frac{1}{2}]$$

then use $a^{[2]}$ and $b^{[2]}$ to get

$$a^{[3]} = [-1, 2, 1, 4, -3, -2, 0, 1].$$

This gives

$$f(x) = -\phi(2^3x) + 2\phi(2^3x-1) + \phi(2^3x-2)$$
$$+4\phi(2^3x-3) - 3\phi(2^3x-4) - 2\phi(2^3x-5) + \phi(2^3x-7).$$

10. Applications of Wavelets

10.1. **Daubechies wavelets.** The most desired basis of a Hilbert space is of course an orthonormal basis. One example of its convenience is that if a scaling function ϕ with two-scale symbol p_k generates a basis of L^2 then the corresponding two-scale symbol q_k of ψ can be found easily by conjugation, sign change and unit shifts. One family of wavelets that can be constructed in this way are the Daubechies wavelets. They are compactly supported and, apart from $\psi_{D1} = \psi_H$, they are continuous. They also have increasing number of continuous derivatives.
To describe how to construct the second Daubechies wavelet, ψ_{D2}, we will start with some theorems. Proofs of the theorems and a more detailed description of the construction can be found in [1]. More information about orthogonal wavelets can be found in [3].

Theorem 10.1. *Take the polynomial $P(x) = \frac{1}{2}\sum_k p_k x^k$ satisfying*

- $P(1) = 1$
- $|P(x)|^2 + |P(-x)|^2 = 1$ *for* $|x| = 1$
- $|P(e^{it})| > 0$ *for* $|t| \leq \frac{\pi}{2}$.

Let $\phi_{D1} = \phi_H$ and let

$$\phi_n(x) = \sum_k p_k \phi_{n-1}(2x-k).$$

Then the sequence ϕ_n converges to the function ϕ satisfying

$$\int_{\infty}^{\infty} \phi(x-n)\phi(x-m)dx = \delta_{nm}$$

and

$$\phi(x) = \sum_k p_k \phi(2x-k).$$

Theorem 10.2. *Let $\{V_j : j \in \mathbb{Z}\}$ be an MRA with a scaling function*

$$\phi(x) = \sum_k p_k \phi(2x - k).$$

Let W_j be the span $\{\psi(2^j x - k) : k \in \mathbb{Z}\}$ where

$$\psi(x) = \sum_k (-1)^k \overline{p_{1-k}} \phi(2x - k).$$

Then $V_{j+1} = V_j \oplus W_j$ and $\{\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k) : k \in \mathbb{Z}\}$ is an orthonormal basis of W_j.

Example 10.3. *Recall the example 8.5. The Haar wavelet ψ_H is the first one in the Daubechies family. It is the only one in the family that is not continuous.*
The Haar scaling function has $p_0 = p_1 = 1$ which, by the means of

$$\psi(x) = \sum_k (-1)^k \overline{p_{1-k}} \phi(2x - k)$$

gives

$$\psi(x) = \phi(2x) - \phi(2x - 1).$$

The problem now is to find the p_j's for the Daubechies wavelets of higher order that satisfies conditions in Theorem 10.1. As mentioned above a suitable procedure for this is described in [1]. For the second Daubechies they are given by

$$p_0 = \frac{1+\sqrt{3}}{4}, p_1 = \frac{3+\sqrt{3}}{4}, p_2 = \frac{3-\sqrt{3}}{4}, p_3 = \frac{1-\sqrt{3}}{4}$$

so that

$$\phi_{D2} = \frac{1+\sqrt{3}}{4}\phi(2x) + \frac{3+\sqrt{3}}{4}\phi(2x-1) + \frac{3-\sqrt{3}}{4}\phi(2x-2) + \frac{1-\sqrt{3}}{4}\phi(2x-3)$$

is the scaling function and

$$\psi_{D2} = \frac{1-\sqrt{3}}{4}\phi(2x) - \frac{3-\sqrt{3}}{4}\phi(2x-1) + \frac{3+\sqrt{3}}{4}\phi(2x-2) - \frac{1+\sqrt{3}}{4}\phi(2x-3)$$

is the wavelet.

For large N the wavelet ψ_{DN} and scaling function ϕ_{DN} have approximately $N/5$ continuous derivatives. Thanks to this and to the compactness of their support, the Daubechies wavelets are very useful in, for example, noise removing, singularity detection and compression.

10.2. **Denoising and compression of images.** Application of wavelets to an image can have two purposes. Either one wants to get a clearer image by getting rid of unwanted noise or to make the data volume of the image smaller by compressing it.

Both for denoising and compression the procedure is roughly done in three steps

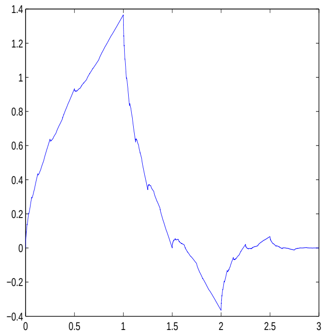

FIGURE 3. The Daubechies wavelet ψ_{D2}.

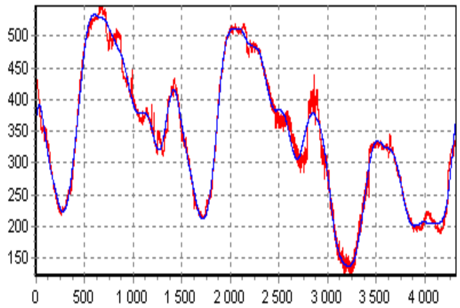

FIGURE 4. Function and its denoised version.

(1) Decompose. A wavelet and a level N is chosen, then the decomposition is computed at that level N.
(2) Selection using thresholding. For each level from 1 to N a threshold is decided and all wavelet coefficients outside that threshold are removed.
(3) Reconstruction. The reconstruction is computed with the original coefficients from level N and the modified coefficients on the levels 1 to N.

The difference between the denoising and the compression is the threshold. If the data of the image is to be compressed, only large enough coefficients are kept. But to denoise the threshold has to be chosen so that the data kept is smoother than before. In other words: coefficients corresponding with too large frequencies are removed. For an image it is still important not to loose all edges. Therefore in the areas of the image with a lot of edges the denoising must be done more carefully.

In Figure 4 a signal with a lot of noise is shown together with the much smoother denoised version of the same signal. All important information is kept but the result looks much nicer.

Applying the same principle to an image gives the same result. Look at Figure 5: The picture to the right is the denoised version of the left picture.

FIGURE 5. Noisy image and denoised version.

FIGURE 6. The original image.

FIGURE 7. The image compressed using threshold 2. The compression ratio is 27:1

For the following compressed images (Figure 7 and 8) the threshold means that any coefficient smaller than the threshold value was taken away. To see the whole procedure and its matlab code, see [11]

When FBI was going to digitalize all fingerprints they used an algorithm based on Daubechies biorthogonal spline wavelets called WSQ, or *The wavelet scalar quantization grey-scale fingerprint image compression algorithm.*
Using this algorithm they where able to get a good image of the fingerprint by only keeping 8% of the data. For more information see [10].

FIGURE 8. The image compressed using threshold 8. The compression ratio is 5:1.

REFERENCES

[1] Boggess, A. and Narcowich F.J., *A First Course in Wavelets with Fourier Analysis.*New Jersey: Prentice-Hall (2001)

[2] Chan, A.K. and Goswami, J.K., *Fundamentals of Wavelets.* New York: Wiley (2005)

[3] Chui, C.K., *An Introduction to Wavelets.* San Diego: Academic Press (1992)

[4] Daubechies, I., *Ten lectures on Wavelets.* Philadelphia: Siam (1992)

[5] Ha, D.M., *Functional Analysis, A Gentle Introduction.* New York: Matrix Editions (2006)

[6] Hellström, L., Morander, S. and Tengstrand, A., *Envariabelanalys.* Lund: Studentlitteratur (1991)

[7] Kreyszig, E., *Introductory Functional Analysis with Applications.* New York: Wiley (1978)

[8] Mallat, S., *A Wavelet Tour of signal Processing.* San Diego: Academic Press (1998)

[9] Rudin, W., *Real and Complex Analysis, 3rd edition.* McGraw Hill (1986)

[10] www.nist.gov *The homepage of NIST information Technology Laboratory.* 2007-11-27

[11] www.owlnet.rice.edu 2007-12-05