CHAPTER-1
OVERVIEW OF WAVELETS

Summary: Recent years have received much attention to wavelets because of its comprehensive mathematical power and good application potential in many interesting physical phenomena. In this introductory chapter, we introduce definition and brief historical development of wavelets, the fundamentals of Hilbert-space, Fourier transform and general theorems needed.

CONTENTS

1. Wavelet, History and Applications
2. Preliminaries
   2.1 Basic Linear Algebra and Hilbert Space
   2.2 Fourier Transform and Properties
   2.3 Discrete Signal
   2.4 Uncertainty Principle
3. Mathematical Theory of Wavelet
   3.1 Continuous Wavelet Transform
   3.2 CWT as an Operator
   3.3 Wavelet Series
4. Multiresolution Analysis
5. General Theorems
6. Generalized Moments

♦ This chapter is based on Sections 1-6 of our publication Int. J. Math. Archive 2(2011), 1776-1793.
1. **Wavelet, History and Applications**

A wavelet is a wave pattern of small size, that is, its graph oscillates only over the short distance or damps very fast; it means the value over the whole domain equates to zero. A wavelet is localizable both in time (position) and frequency (scale).

**Wavelet** [30]: An oscillatory function \( \psi(x) \in L^2(R) \) with zero mean is a wavelet if it has the desirable properties:

1. **Smoothness**: \( \psi(x) \) is \( n \) times differentiable and that their derivatives are continuous.

2. **Localization**: \( \psi(x) \) is well localized both in time and frequency domains, i.e. \( \psi(x) \) and its derivatives must decay very rapidly. For frequency localization \( \hat{\psi}(\omega) \) must decay sufficiently fast as \( |\omega| \to \infty \) and that \( \hat{\psi}(\omega) \) becomes flat in the neighborhood of \( \omega = 0 \). The flatness is associated with number of vanishing moments of \( \psi(x) \), i.e.

\[
\int_{-\infty}^{\infty} x^k \psi(x) dk = 0 \quad \text{or equivalently} \quad \frac{d^k \hat{\psi}(\omega)}{d\omega^k} = 0 \quad \text{for} \quad k = 0, 1, \ldots, n.
\]

in the sense that larger the number of vanishing moments more is the flatness when \( \omega \) is small.

3. **The admissibility condition**

\[
\int_{-\infty}^{\infty} \left| \frac{\hat{\psi}(\omega)}{\omega} \right|^2 d\omega < \infty
\]

suggests that \( |\hat{\psi}(\omega)|^2 \) decays at least as \( |\omega|^{-1} \) or \( |x|^{\varepsilon - 1} \) for \( \varepsilon > 0 \).

Let \( \psi_{a,b}(x), \ a \in R, b \in R \) be a family of functions generated from mother wavelet \( \psi(x) \) by scaling \( (a) \) and translation \( (b) \) and defined by

\[
\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{x-b}{a}\right), \quad \|\psi_{a,b}(x)\| = \|\psi\|_2.
\]

Notice that \( a \) is a measure of degree of compression and \( b \) signifies that \( \psi_{a,b} \) is centred (localized) around \( b \). \( \{ \psi_{a,b}(x) \} \) is an orthonormal basis of \( L^2(R) \).
Dyadic Wavelet. Let $a = 2^{-j}$ and $b = k2^{-j}$. The function $\psi_{j,k}$ stands for the dyadic wavelet shrunk by a factor of $2^j$ if $j$ is positive (magnified by a factor $2^{-j}$ if $j$ is negative) and shifted by $k2^{-j}$ units.

Examples of Wavelets ([30] & [89, pp. 288-289])

1. Gaussian wavelet:
   \[ \psi(x) = cxe^{-\pi^2} \]

2. Mexican Hat or Maar’s Wavelet:
   \[ \psi(x) = \frac{d}{dx} \left( \frac{cxe^{-x^2}}{2\pi} \right) = c \left( \frac{1}{2\pi} - x^2 \right)e^{-\pi^2} \]

3. Haar Wavelet (1910):
   \[ \psi(x) = \begin{cases} 
   1 & 0 \leq x < \frac{1}{2} \\
   -1 & \frac{1}{2} \leq x < 1 \\
   0 & \text{otherwise} 
   \end{cases} \]

4. Poisson Wavelet:
   \[ \psi(x) = -\left(1 + \frac{d}{dx} \right) \frac{1}{\pi} \frac{1}{1+x^2} \]

5. Morlet wavelet:
   \[ \psi(x) = \exp \left( i\omega x - \frac{x^2}{2} \right), \]

The word, wavelet (or ondelet) was first introduced by J. Morlet, a French Geophysicist working in an oil company, Elf. Acquitaine, at the beginning of 1980’s. The wavelet, which started attracting the scientific community of early eighties, is a synthesis of ideas originated from various specialties including Mathematics (Harmonic analysis: Calderon-Zygmund Operator, Little Wood-Paley Theory (1937), Franklin basis, and atomic decomposition of function), Physics (coherent states formalism in quantum mechanics and renormalization
group) and Engineering (quadratic mirror filters side bent coding in signal processing and pyramidal algorithm in image processing).

Indeed the first orthonormal wavelet basis was discovered by Alfred Haar (1909) and was refined by P. Franklin in 1927. J.O. Stromberg (1981) was of course the first to be credited for constructing orthonormal basis of \( L^2(R) \): \( \psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k), j, k \in \mathbb{Z} \), where \( j, k \) represent scale and translation parameters respectively. For positive \( j \) the graphical display of \( \psi_{j,k} \) is wider and flatter, whereas for negative \( j \), the same is narrower and sharper.

In 1982, Morlet introduced the idea of transform and, in 1984, Grossman and Morlet succeeded in establishing the inversion formula. In 1985, Y. Meyer, a pure mathematician, used Littlewood Paley methods of 1930’s and Calderon’s method of 1960’s to formalize the notion of a wavelet. In 1986, Mallat realized that coarse features in an image are large objects, whereas fine scale feature should be studied much more locally. Subsequently, Daubechies, Grossman, Mallat, Meyer and Strang have developed the theory of wavelets to a considerable extent. The first application of a wavelet is due to Morlet (1983) and is

\[
\psi(x) = \pi^{-3/4} \int_{\mathbb{R}} e^{ikx} e^{-e^{-k^2/2}} \, dk \quad \text{at} \quad k = \pi(2 \log 2)^{1/2}.
\]

Wavelet analysis is probably the most recent solution to overcome the shortcoming of Fourier transform. In the case of wavelet, we normally do not speak about time-frequency representation but about time–scale representation, scale being in a way the opposite of frequency, because the term frequency is reserved for the Fourier transform, since from literature it is not always clear what is meant by small and large scales. It is defined as follows: The large scale is the big picture, while the small scale shows the details. Thus going from large to small scale is in this context equal to zooming in.
Applications of Wavelets

Electronics (signal compression and denoising; image and speech analysis), Computer (computer graphics, neural network), Mathematics (approximation theory, matrix theory, numerical analysis of ODEs and PDEs, operator theory, inverse problems), Mathematical Statistics (sampling theory, regression, density and function estimation, factor analysis modeling and forecasting in time series analysis, spatial statistics, pattern recognition), Meteorology (structure of the clouds), Universe (structure of galaxies and universe), Biomedical (bio-acoustics, electro-cardiography (ECG), electroencephalography (EEG)), Biomedical Imaging (biomedical image processing, i.e. noise reduction, image enhancement and detection of micro calcification in mamograms, computer assisted magnetic resonance imaging (MRI), functional image analysis), Fluid (turbulence), Mathematical finance and many more.

For basic theory of wavelet and its applications, see Altaisky [1], Chakraborty et al. [20], Debnath [30], Gannam [52], Hamrita et al. [55], Holschneider [60], Islam [61], Jensen et al. [63], Nagendra et al. [82], Nasif et al. [85], Nielsen [86], Pinsky [89], Rao et al. [94], Resnikoff et al. [100], Ruikar et al. [102], Sadashivappa et al. [104], Sarkar et al. [105], Schumaker et al. [106], Soman et al. [110], Vidakovic [114], Vuorenmaa [115], Wadi et al. [116-117], Walker [118] and Zayed [125].

2. Preliminaries

2.1 Basic Linear algebra and Hilbert Space ([25], [30], [46] & [89])

Definition 1. Let $H$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$. A set of vectors $\{x_n\}$ is an orthonormal system if $\langle x_n, x_m \rangle = \delta_{mn}$.

Lemma 1. A set of vectors $\{x_n\}$ is orthonormal iff for every finite set of complex numbers $\{a_n\}$, we have
\[
\left\| \sum_n a_n x_n \right\|^2 = \sum |a_n|^2.
\]

**Definition 2.** Let \( H \) be a Hilbert space. A set of vectors \( \{x_n\} \) is a Riesz system, if there exist constants \( 0 \leq c \leq C < \infty \) such that for any finite set of complex numbers \( \{a_n\} \)

\[
c \left\| \sum_n a_n x_n \right\|^2 \leq \left\| \sum_n a_n x_n \right\|^2 \leq C \sum_n |a_n|^2.
\]

**Definition 3.** Let \( L^2(R) \) is a vector space of square integrable function, i.e.

\[
L^2(R) : \left\{ f : R \to C : \int |f(x)|^2 dx < \infty \right\}.
\]

For \( f, g \in L^2(R) \), define inner product \( \langle f, g \rangle = \int f(x) \overline{g(x)} dx \).

In particular \( \|f\|_2 = \left( \int |f(x)|^2 dx \right)^{1/2} \), and we say that \( f \) is square integrable.

**Definition 4.** Let \( L^1(R) : \left\{ f : R \to C : \int |f(x)| dx < \infty \right\} \). For \( f, g \in L^1(R) \), let \( \|f\|_1 = \int |f(x)| dx \).

We say that \( f \) is integrable.

**Lemma 2.** If \( f \in L^1(R) \), then \( \left\| \int f(x) dx \right\| \leq \int |f(x)| dx = \|f\|_1 \).

**Definition 5.** Let \( f : R \to C \) be a function. The support of \( f \), denoted \( \text{supp} f \), is the closure of the set \( \{x \in R : f(x) \neq 0\} \). We say \( f \) has compact support if \( \text{supp} f \) is a compact set.

In other words, \( f \) has compact support if there exists \( r < \infty \) such that \( \text{supp} f \subseteq [-r, r] \), that is, such that \( f(x) = 0 \) for all \( x \) satisfying \( |x| > r \).

**Definition 6.** A sequence \( \{\varphi_n\} \) of orthonormal basis in a Hilbert space \( H \) is called a frame if there exist constants \( A, B > 0 \) such that
\[ A\|f\|^2 \leq \sum_{n=1}^{\infty} |(f, \varphi_n)|^2 \leq B\|f\|^2 \quad \forall \ f \in H \]

The constants \( A \) and \( B \) are called frame bounds. If \( A = B \), then frame is called tight.

**Cauchy-Schwartz Inequality.** For \( f, g \in L^2(R) \),

\[
\left| \int f(x) \overline{g(x)} \, dx \right| \leq \left( \int |f(x)|^2 \, dx \right)^{1/2} \left( \int |g(x)|^2 \, dx \right)^{1/2}, \text{ i.e.} \]

\[
|\langle f, g \rangle| \leq \|f\| \|g\|. \]

**Triangle Inequality.** For \( f, g \in L^2(R) \),

\[
\left( \int |f(x) + g(x)|^2 \, dx \right)^{1/2} \leq \left( \int |f(x)|^2 \, dx \right)^{1/2} + \left( \int |g(x)|^2 \, dx \right)^{1/2}, \text{ i.e.} \]

\[
\|f + g\| \leq \|f\| + \|g\|. \]

**Lemma 3.** Suppose \( f, g \in L^2(R) \), satisfies a Lipschitz condition of order \( \alpha \in (0,1] \), which means that there exists a constant \( c < \infty \) such that for all \( x, y \in R \)

\[
|f(x) - f(y)| \leq c|x - y|^{\alpha}. \]

**2.2 Fourier Transform and Properties**

**Fourier Transform** [30]. Fourier transform [FT] is a well known mathematical tool to transform time domain signal to frequency domain for efficient extraction of information and vice- versa.

**Definition 7 (Fourier Transform).** For \( f \in L^1(R) \) or \( L^2(R) \) and \( \omega \in R \), define
\[ F(f) = \hat{f}(\omega) = \int_{\mathbb{R}} f(x)e^{-ix\omega} \, dx, \]

where \( \hat{f} \) is called the FT of \( f \) and the mapping \( \hat{\cdot} \) is called the FT.

**Inverse FT.** For \( g \in L^1(\mathbb{R}) \) and \( x \in \mathbb{R} \), we define \( g^\vee \), the Inverse FT of \( g \) by

\[ g^\vee(x) = \int_{\mathbb{R}} \hat{g}(\omega)e^{ix\omega} \, d\omega. \]

The mapping \( ^\vee \) is the Inverse FT.

**Characteristics of FT** ([16, pp. 101-110] & [114, pp. 30-31])

Boundedness: \( \hat{f} \in L^\infty(\mathbb{R}) \), \( \| \hat{f} \|_\infty \leq \| f \|_1 \)

Uniform Continuity: \( \hat{f}(\omega) \) is uniformly continuous on \( -\infty < \omega < \infty \).

Decay: For \( f \in L^1(\mathbb{R}) \), \( \hat{f}(\omega) \to 0 \) when \( |\omega| \to \infty \) [Reimann Lebesgue Lemma].

Linearity: \( F[\alpha f(x) + \beta g(x)] = \alpha F[f(x)] + \beta F[g(x)] \).

Derivative: \( F[f^{(n)}(x)] = (i\omega)^n \hat{f}(\omega) \).

Plancherel’s Identity: \( \langle f, g \rangle = \langle \hat{f}, \hat{g} \rangle \). If \( g = f \), then the above identity reduces to

\[ \| f \|^2 = \| \hat{f} \|^2. \]

The function \( |\hat{f}(\omega)|^2 \) is called the energy spectrum. Analogously, the area below the curve \( |\hat{f}(\omega)|^2 \) is equal to \( \int |f(x)|^2 \, dx \) - the energy content of the signal.

Shifting: \( Ff(x-x_0) = e^{-i\omega x_0} \hat{f}(\omega) \).
Scaling: \( Ff(ax) = \frac{1}{|a|} f\left(\frac{\omega}{a}\right) \).

Symmetry: \( F[F[f(x)]] = f(-x) \).

Convolution: The convolution of \( f \) and \( g \) is defined as
\[
f^* g(x) = \int f(x-t)g(t)dt. \quad F \left[ f^* g(x) \right] = \hat{f}(\omega)\hat{g}(\omega).
\]

Modulation Theorem: \( f(x)g(x) = F(\omega) \ast G(\omega) \) [by symmetry property].

Moment Theorem: \( \int x^n f(x)dx = (i)^n \frac{d^n \hat{f}(\omega)}{d\omega^n} \)

2.3 Discrete Signal [7, p. 385]

For some function \( f \)
\[
f_j = F\left(\frac{j}{M}\right), \quad j = 0,1, \ldots, M - 1.
\]

Fourier coefficients \( \hat{f}_n = \frac{1}{M} \sum f_j e^{-2\pi ij/M} \), \( n = \frac{M}{2} + 1, \ldots, 0, \ldots, -\frac{M}{2} \).

2.4 Uncertainty Principle [16, pp. 123-127]

**Definition 8.** Let \( f \in L^2(R) \). The dispersion of \( f \) about the point \( a \in \mathbb{R} \) is the quantity
\[
\Delta_a f = \frac{\int_{-\infty}^{\infty} (x-a)^2 |f(x)|^2 dx}{\int_{-\infty}^{\infty} \int |f(x)|^2 dx}.
\]
The dispersion about a point $a$ is the measure of deviation or spread of its graph from $x = a$. This dispersion will be small if the graph of $f$ is concentrated near $t = a$ and is spread out away from $x = a$.

In frequency domain,

$$
\Delta_a \hat{f} = \frac{\int_{-\infty}^{\infty} (\omega - a)^2 |\hat{f}(\omega)|^2 d\omega}{\int_{-\infty}^{\infty} |\hat{f}(\omega)|^2 d\omega}.
$$

**Theorem 1 (Uncertainty Principle).** Suppose $f$ is a function in $L^2(R)$ which vanish at $+\infty$ and $-\infty$. Then

$$
\Delta_a f \cdot \Delta_a \hat{f} \geq \frac{1}{4}
$$

for all points $a, \alpha \in \mathbb{R}$.

The statement implies that $\Delta_a f$ and $\Delta_a \hat{f}$ cannot simultaneously be small. In other words, when the time-frequency cell is narrow in time it is wider in frequency and vice-versa. In case of Gaussian function $f(x) = \frac{1}{\sqrt{2\pi \sigma}} e^{-\frac{x^2}{2\sigma^2}}$ equality is achieved.

3. **Mathematical Theory of Wavelet**

3.1 **Continuous Wavelet Transform [CWT]**

The wavelet transform (continuous or discrete) or wavelet analysis is probably the most recent solution to overcome the shortcomings of the FT. The CWT can effectively treat signals or function $f(x)$ with spikes whose Fourier series would require many high-frequency components [7, p. 480].

Wavelet constitutes a family of functions derived from one single function and indexed by two labels, one for position and other for frequency. That is, the wavelet transform of a one dimensional function is two dimensional; the wavelet transform of a two dimensional function is four dimensional. One imposes some additional condition on the wavelet function
in order to make the wavelet transform decrease quickly with decreasing scale. These are the regularity conditions and state that the wavelet function should have some smoothness and concentration in both time and frequency domains.

The CWT of a function \( f(x) \in L^2(\mathbb{R}) \) at a scale \( a \) and position \( b \) with respect to \( \psi(x) \in L^2(\mathbb{R}) \) is given by ([1, p. 10], [30] & [94, p. 17])

\[
W_\psi(f)(a,b) = <f, \psi_{a,b}> = \int_{-\infty}^{\infty} f(x) \overline{\psi_{a,b}(x)} dx = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(x) \psi\left(\frac{x-b}{a}\right) dx .
\]

In \( \psi_{a,b} \) the parameter \( b \) gives the position of the wavelet, while the dilation parameter \( a \) governs frequency. For smaller values of \( a (>0) \), the wavelet is contracted in the time domain and the wavelet transform gives information about the finer details of the signal; while for large values of \( a \), the wavelet expands and the wavelet transform gives a global view of signal. Fig.1 [94, p.19] shows two dilations of the Morlet wavelet. If \( a > 1 \) there is a stretching of \( \psi(x) \) along the time axis whereas if \( 0 < a < 1 \) there is a contraction of \( \psi(x) \).

Fig.1: A Morlet wavelet dilated by factor of \( a = 1/2 \) and \( a = 3 \).
If \( a \) and \( b \) assume only the discrete values, then the corresponding wavelet transform will be Discrete wavelet transform [DWT.]

**Inverse CWT** [30]

\( f(x) \) can be recovered from its CWT as

\[
f(x) = \frac{1}{C} \int_{a=0}^{\infty} \int_{b=-\infty}^{\infty} \frac{1}{|\alpha|^2} W(a, b) \psi_{a, b} \, da \, db,
\]

\[
C = \int_{0}^{\infty} \left| \hat{\psi}(\omega) \right|^2 \, d\omega.
\]

### 3.2 The CWT as an Operator

The CWT takes a member of the set of square integrable function of one real variable in \( L^2(R) \) and transforms it to a member of the set of functions of two real variables. Thus, it can been seen as a mapping operator from \( L^2(R) \) to the latter set.

Define \( W_\psi[f(x)] \equiv W(a, b) \). Then \( W_\psi[f] \) is to be read CWT with respect to \( \psi(x) \) of \( f \). The notation for the operator use \( \psi \) as a subscript to remind us of the fact that the transform depends not only on the function \( f(x) \) but also on the mother wavelet.

We now enumerate various properties of CWT using the operator notation ([1, p. 12], [30] & [94, pp. 33-34]):

**Linearity:**

\[
W_\psi[\alpha f(x) + \beta g(x)] = \alpha W_\psi[f(x)] + \beta W_\psi[g(x)]
\]

for scalar \( \alpha, \beta \) and function \( f(x), g(x) \in L^2(R) \).

**Translation:**

\[
W_\psi[f(x - \tau)] = W[a, b - \tau]
\]
Scaling:

\[ W_\phi \left[ \frac{1}{\sqrt{\alpha}} f \left( \frac{1}{\alpha} \right) \right] = W \left[ \frac{a}{\alpha}, \frac{b}{\alpha} \right] \quad \text{for } \alpha > 0 \]

**Wavelet Shifting:** Let \( \hat{\psi}(x) = \psi(x - \tau) \). Then

\[ W_\psi [f (x)] = W(a, b + a \tau) . \]

Observe that the CWT obtained by shifting the wavelet is different from the one obtained by shifting the signal.

**Energy Conservation:**

\[ \int_{-\infty}^{\infty} |f(x)|^2 \, dx = \frac{1}{C} \int_{-\infty}^{\infty} \int_{0}^{\infty} |< f, \psi_{a,b} >|^2 \, \frac{dadb}{a^2} . \]

**Localization:** Let \( f(x) = f(x - x_0) \) be the Dirac pulse at the point \( x_0 \), then

\[ W_\psi [f](a, b) = \frac{1}{\sqrt{a}} \psi \left( \frac{x_0 - b}{a} \right) . \]

**3.3 Wavelet Series**

A function \( \psi \in L^2 (R) \) is said to be orthonormal wavelet if the family \( \{ \psi_{j,k} \}_{j,k \in \mathbb{Z}} \), where

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

satisfies the conditions

\[ < \psi_{j,k}, \psi_{l,m} >= \delta_{j,l} \delta_{k,m}; \quad j, k, l, m \in \mathbb{Z} \]

Wavelet series expansion of \( f \in L^2 (R) \) is defined by

\[ f(x) = \sum_{j,k=-\infty}^{\infty} \beta_{j,k} \psi_{j,k}(x) , \]

where the wavelet coefficients

\[ \beta_{j,k} = < f, \psi_{j,k} > = \int_{-\infty}^{\infty} f(x) \overline{\psi}_{j,k}(x) \, dx . \]
4. Multiresolution Analysis

The purpose of multiresolution analysis is to write a function \( f \in L^2(R) \) as a collection (sequence) of its successive approximations, each of which is a smoothed version of the previous one.

**Definition 9.** A multiresolution analysis [MRA] of \( L^2(R) \) is a sequence \( \{V_n\}_{n \in \mathbb{Z}} \) of the closed subspaces of functions \( f \in L^2(R) \) satisfying the following properties ([7, pp. 414-418], [16, p. 191],[30], [89, p.304] & [125]):

(i) (Monotonicity) is \( V_n \subset V_{n+1} \quad \forall \ n \in \mathbb{Z} \) \( \text{ and } V_n \subset V_{n-1} \quad \forall \ n \in \mathbb{Z} \)

(ii) (Separation) \( \bigcap_{n=-\infty}^{\infty} V_n = \{0\}, \quad \bigcup_{n=-\infty}^{\infty} V_n \) is dense in \( L^2(R) \) i.e. \( \bigcup_{n=-\infty}^{\infty} V_n = L^2(R) \)

(iii) (Dilation) \( f(x) \in V_0 \) iff \( f(2^n x) \in V_n \forall n \in \mathbb{Z} \), i.e. all the spaces are scaled versions of the central space \( V_0 \).

(iv) (Existence of scaling) There exists a scaling function \( \varphi(x) \in V_0 \) whose integer translates space \( V_0 \), i.e. for each \( f(x) \in L^2(R) \)

\[
V_0 = \left\{ f(x) = \sum_{k=-\infty}^{\infty} c_k \varphi(x-k) \right\}
\]

and \( \{\varphi(x-k), k \in \mathbb{Z}\} \) is an orthonormal basis for \( V_0 \).

The function \( \varphi \in V_0 \) is called scaling function.

Suppose \( \varphi_{j,k}(x) = 2^{j/2} \varphi(2^j x - k), j, k \in \mathbb{Z}, x \in R \). Since \( \varphi_{0,k}(x) \in V_0 \forall k \in \mathbb{Z} \) due to (iv).

Further, if \( j \in \mathbb{Z} \) condition (iii) implies that the family \( \{\varphi_{j,k}(x) = 2^{j/2} \varphi(2^j x - k), j, k \in \mathbb{Z}\}_{k=-\infty}^{\infty} \) is an orthonormal basis for \( V_j \).

By definition part (iv) means that for any \( f \in L^2(R) \), there exists a sequence \( \{f_n\}_{n=1}^{\infty} \) such that each \( f_n \in \bigcup_{j \in \mathbb{Z}} V_j \) and \( \{f_n\}_{n=1}^{\infty} \) converges to \( f \) in \( L^2(R) \), that is \( \|f_n - f\| \to 0 \) as \( n \to \infty \).

The functions, consisting of translations and dilations of wavelet function \( \psi(2^j x - k) \), form a complete and orthonormal basis of \( L^2(R) \).

The relation between two functions is expressed as
\[ V_{j+1} = V_j \oplus W_j \quad W_j \perp W_k \text{ for } j \neq k \]

where subspaces \( V_j = 2^{j/2} \varphi(2^j x - k); \ k = ... , -1, 0, 1, ... \)

\[ W_j = 2^{j/2} \psi(2^j x - k); \ k = ... , -1, 0, 1, ... \]

Thus for \( k > 0 \), \( V_{j+1} = V_0 \oplus (\bigoplus_{j=0}^{j} W_j) \), i.e. \( V_{j+1} \) can be expressed as a linear combination of functions in orthogonal spaces \( V_0 \) and \( W_j \), \( j = 0, 1, ... , j \) and analysed separately at different scales.

Since \( \bigcup_{j=-\infty}^{\infty} V_j = L^2(R) \). For \( j \to \infty \), \( V_0 + (\bigoplus_{j=0}^{\infty} W_j) = L^2(R) \).

Similarly \( V_0 = V_{-1} \oplus W_{-1} = V_{-j} \odot W_{-j} \). But \( \bigcap_{j=-\infty}^{\infty} V_j = \{0\} \) implies \( V_{-j} \to \{0\} \) as \( j \to \infty \). We get \( (\bigoplus_{j=-\infty}^{\infty} W_j) = L^2(R) \). Therefore, \( W_j \) is a decomposition of \( L^2(R) \) into mutually orthogonal subspaces. Thus, \( \psi(x) \in W_0 \) such that \( \psi_{j,k} \) is a complete orthonormal basis of \( W_j \), i.e. \( \{\psi_{j,k}(x)\} \) is an orthonormal basis of \( L^2(R) \).

Note that for certain values of \( j \) and \( N \),

\[ \text{supp } \varphi_{j,k} = \left[ \frac{k}{2^j}, \frac{N+k-1}{2^j} \right] \]

\[ \text{supp } \psi_{j,k} = \left[ \frac{k-N/2}{2^j}, \frac{k+N/2}{2^j} \right] \]

A scaling function can be used to expand a general function: Projection \( P_j : L^2(R) \xrightarrow{onto} V_j \),

\[ f(x) = P_j f = \sum_{k \in \mathbb{Z}} 2^{j/2} c_k \varphi(2^j x - k). \]

It satisfies the following convergence property

\[ \left\| f - \sum_{k} c_k \varphi(2^j x - k) \right\| \leq C2^{-jp} \| f^{(p)} \| \text{ for large } j, \]

where \( c_k = \int f(x) \varphi(2^j x - k) dx \); \( C, p \) are constants.

Moment = 0 implies scaling bases can be represented as polynomials of degree \( (\frac{N}{2} - 1) \).
Theorem 2 [16, p. 205]. Let $V_j$, $j \in \mathbb{Z}$ be a given MRA with scaling function $\varphi$ and $P_j f$, projection of $f \in L^2(R)$ onto $V_j$ such that
\[
P_j f = \sum_{k \in \mathbb{Z}} 2^{j/2} c_k \varphi(2^j x - k), \text{ where } c_k = \int f(x) \varphi(2^j x - k) dx
\]
then for $j$ sufficiently large
\[
c_k \cong m f(k2^{-j}) \text{ with } m = \int \varphi(x) dx.
\]

Remarks [46, pp. 380-381]. Let for $r > 0$, $\psi(x)$ has compact support, that is $\psi(x) = 0 \ \forall x \ s.t \ |x| > r$. This means $\psi(2^j x)$ has compact support inside the interval $[-\frac{r}{2^j}, \frac{r}{2^j}]$ since $\psi(2^j x) = 0$, whenever $|2^j x| > r$, i.e. when $|x| > \frac{r}{2^j}$.

The graph of $\psi(2^j x - k) = \psi(2^j (x - 2^{-j} k))$ is obtained by translating the graph of $\psi(2^j x)$ by $2^{-j} k$ along x-axis (to the right if $k > 0$ and to the left if $k < 0$). Hence, if compact support of $\psi$ is in $[-r, r]$, then $\psi(2^j x - k)$ has support inside $[2^{-j} k - 2^{-j} r, 2^{-j} k + 2^{-j} r]$. Finally, the graph of $\psi_{j,k}$ is obtained from the graph of $\psi(2^j x - k)$ by after multiplication by $2^{j/2}$, which stretches the graph in y direction by this factor. For $r$ very small $\psi_{j,k}$ is centered near the point $2^{-j} k$ and has a scale of about $2^{-j}$.

5. General Theorems ([7], [16], [30], [46], [54], [86], [110] & [125])

If $\{\varphi_n\}$ is an orthonormal basis then it is a tight frame, since
\[
\sum_{n=1}^{\infty} |< f, \varphi_n >|^2 = \|f\|^2.
\]

Sufficient condition for $\{\psi_m\}$ to constitute a frame in $L^2(R)$.

Theorem 3. Let $\psi$ and $a_0$ such that
\[
i \inf_{|x| \leq a_0} \sum_{m=-\infty}^{\infty} |\psi(a_0^n \omega)|^2 > 0
\]
ii) \( \sup_{|\omega| \leq a_0} \sum_{m=\pm \infty} |\hat{\psi}(a_0^m \omega)|^2 > 0 \)

iii) \(|\hat{\psi}(a_0^m \omega + x)| \leq c(1 + |x|)^{-\epsilon(1+\epsilon)} \) for some \( \epsilon > 0 \), i.e. decays at least as fast as \((1 + |x|)^{-\epsilon(1+\epsilon)}\) for some \( \epsilon > 0 \),

then there exists \( \bar{b} > 0 \) such that \( \psi_{mn}(x) = a_0^{m/2} \psi(a_0^m x - nb_0) \) form a frame for any \( b_0 < \bar{b}_0 \), i.e. for any \( b_0 \in (0, \bar{b}) \).

**Theorem 4.** For the scaling function it holds \( \int_{\mathbb{R}} \phi(x) \, dx = 1 \) or equivalently \( \hat{\phi}(0) = 1 \), where \( \hat{\phi}(\omega) \) is the FT of \( \phi \).

**Theorem 5.** For a given multiresolution analysis, there exists an orthonormal wavelet basis for \( L^2(\mathbb{R}) \). Let \( W_j \) be the orthogonal complement of \( V_j \) in \( V_{j+1} \) Then \( L^2(\mathbb{R}) = V_0 + (\bigoplus_{j=0}^{\infty} W_j) \). In particular, each \( f \in L^2(\mathbb{R}) \) can be uniquely expressed as a sum \( \sum_k w_k \) with \( w_k \in W_k \), where \( w_k \)'s are mutually orthogonal. Equivalently, the set of all wavelets \( \{\psi_{j,k}\}_{j,k \in \mathbb{Z}} \) is an orthogonal basis for \( L^2(\mathbb{R}) \).

**Theorem 6.** For any \( n \in \mathbb{N} \) there exists Daubechies MRA with function \( \phi \) and \( \psi \) that have compact support of length \( 2N - 1 \). Moreover, Daubechies wavelet has \( N \) vanishing moments, i.e.

\[ \langle x^k, \psi \rangle_{L^2(\mathbb{R})} = 0, \quad k = 0, N - 1. \]

**Theorem 7.** Let \( \psi \) be an admissible mother wavelet satisfying

\[ \int (1 + |x|) |\psi(x)| \, dx < \infty. \]

i) If \( f \) is bounded function that satisfies Lipschitz condition of order \( \alpha \), \( 0 < \alpha \leq 1 \), then

\[ |W \psi f(a, b)| \leq c |a|^\alpha^{1/2} \] for some constant \( c > 0 \).
ii) If \( f \) is bounded and continuous at \( x_0 \) with \( 0 < \alpha \leq 1 \), i.e.
\[ |f(x + \lambda) - f(x_0)| \leq \alpha |\lambda|^\alpha \quad \text{for some } \alpha > 0. \]

Then
\[
\left| W_{\varphi} f(a, b) \right| \leq c |a|^{1/2}(|a|^2 + |b|^2), \text{ for some constant } c > 0.
\]

**Lemma 4.** Suppose \( \varphi \in L^1(R) \cap L^2(R) \) satisfies \( \int_{-\infty}^{\infty} \varphi(x) \, dx = 1 \) and
\[ \int |x|^\alpha |\varphi(x)| \, dx = c_2 < \infty. \]

Then \( \varphi \) satisfies:
\[ \left| 2^{j/2} (f, \varphi_{jk}) - f(2^{-j}k) \right| \leq c_1 c_2 2^{-\alpha}.
\]

Or equivalently,
\[ \left| (f, \varphi_{jk}) - 2^{-j/2} f(2^{-j}k) \right| \leq c_1 c_2 2^{-j/2 - (\alpha - 1)/2}. \]

**Theorem 8.** Let \( \varphi \) be continuous function with compact support that satisfies:
i) \( \{ \varphi(x - k) \}_{k \in \mathbb{Z}} \) is an orthonormal system
ii) \( \int_{-\infty}^{\infty} \varphi(x) \, dx = 1 \)
iii) Only finite number of the coefficients \( a_k \) in
\[ \varphi(x) = \sum_k a_k \varphi(2x - k) \]
are non-zero.

Then \( \varphi \) is a scaling function, i.e. \( \varphi \) can be used in construction of MRA.

**Theorem 9.** Suppose that the polynomial \( P(z) = \frac{1}{2} \sum_k a_k z^k \) satisfies:
i) \( P(1) = 1 \)
ii) \( |P(z)| \neq 0 \) for any \( z \) with \( |z| = 1 \)
iii) \( |P(z)|^2 + |P(-z)|^2 = 1 \) for any \( z \) with \( |z| = 1 \)

Then the iteration \( \varphi_0 = \chi_{(0,1)} \), \( \varphi_n(x) = \sum_k a_k \varphi_{n-1}(2x - k) \) converges pointwise and in \( L^2(R) \) to a scaling function \( \varphi \).
Lemma 5. Let $\varphi, \psi \in L^2(R)$, then
i) The set $\{\varphi(x-n)\}_{n=-\infty}^{\infty}$ is orthonormal iff $\sum|\hat{\varphi}(\omega + 2\pi k)|^2 = 1$

ii) The set $\{\varphi(x-n)\}_{n=-\infty}^{\infty}$ and $\{\psi(x-m)\}_{m=-\infty}^{\infty}$ are biorthogonal, i.e.

$\langle \varphi_n, \psi_m \rangle = 0 \ \forall m, n$ iff $\sum \hat{\varphi}(\omega + 2\pi k) \hat{\psi}(\omega + 2\pi k) = 0$,

where $\varphi_n(x) = \varphi(x-n)$ and $\psi_m(x) = \psi(x-m)$.

Lemma 6. The scaling function $\varphi$ satisfies the following conditions:

i) $\sum|\hat{\varphi}(\omega + 2\pi k)|^2 = 1$

ii) $\hat{\varphi}(\omega) = H\left(\frac{\omega}{2}\right)\hat{\varphi}\left(\frac{\omega}{2}\right)$

where $H(\omega)$ is a $2\pi$ periodic function that belongs to $L^2[0,2\pi]$ and satisfies

$|H(\omega)|^2 + |H(\omega + 2\pi)|^2 = 1$.

Lemma 7. The set $\{\psi(x-m)\}_{m=-\infty}^{\infty}$ are orthonormal iff

$\sum|\hat{\psi}(\omega + 2\pi k)|^2 = 1$

and

$\hat{\psi}(\omega) = M\left(\frac{\omega}{2}\right)\hat{\psi}\left(\frac{\omega}{2}\right)$,

where $M(\omega)$ is a $2\pi$ periodic function that belongs to $L^2[0,2\pi]$ and satisfies

$|M(\omega)|^2 + |M(\omega + \pi)|^2 = 1$.

Lemma 8. The FT of any function $f \in W_0$ can be written in the form $\hat{f}(\omega) = \lambda_f(\omega) \hat{\psi}(\omega)$, where $\lambda_f(\omega)$ is a periodic function with period $2\pi$ and $\hat{\psi}$ is independent of $f$. Moreover, $\lambda_f \in L^2[0,2\pi]$ and $\|f\|_{L^2(R)} = \|\lambda_f\|_{L^2[0,2\pi]}$. 
Frequency Domain Characterization of Filter Coefficients

FT of \( \{h(k)\} \) is

\[
H(\omega) = \frac{1}{\sqrt{2}} \sum_k h(k) e^{-i k \omega}.
\]

FT of \( \{g(k)\} \) is

\[
M(\omega) = \frac{1}{\sqrt{2}} \sum_k g(k) e^{-i k \omega}.
\]

Notice that \( H(\pi) = 0 \) and \( H(0) = 1 \).

[\( \hat{\varphi}(\omega) = H\left(\frac{\omega}{2}\right) \hat{\varphi}\left(\frac{\omega}{2}\right) = \hat{\varphi}(0) \prod_{j=1}^{\infty} H\left(\frac{\omega}{2^j}\right), \quad \left[ \varphi(t) = \sum_{k=-\infty}^{\infty} h_k \sqrt{2} \varphi(2t - k) \right] \)]

[\( \hat{\psi}(\omega) = M\left(\frac{\omega}{2}\right) \hat{\psi}\left(\frac{\omega}{2}\right) = M\left(\frac{\omega}{2}\right) \prod_{j=1}^{\infty} M\left(\frac{\omega}{2^j}\right), \quad \left[ \psi(t) = \sum_{k=-\infty}^{\infty} g_k \sqrt{2} \varphi(2t - k) \right] \)]

**Lemma 9.** Suppose \( M: R \to C \) satisfies \( M(0) = 1, |M(\omega)| \leq 1 \) for all \( \omega \in R \) and there exist \( \alpha > 0 \) and \( c < \infty \) such that

\[
|M(\omega)| - |M(0)| \leq c |\omega|^\alpha \quad \forall \omega \in R
\]

For \( n \in \mathbb{N} \), let

\[
M_n(\omega) = \prod_{j=1}^{n} M\left(\frac{\omega}{2^j}\right).
\]

Then \( M_n(\omega) \) converges as \( n \to \infty \) uniformly on every bounded subset of \( R \), hence pointwise at every point \( \omega \in R \).

**Theorem 10.** Let \( p = \frac{N}{2} \) be the number of vanishing moments for a wavelet \( \psi_{jk} \) and let \( f \in C^p(R) \). Then the wavelet coefficients decay as following:

\[
|d_{jk}| \leq C_p 2^{-j\left(p + \frac{1}{2}\right)} \max_{\omega \in I_{jk}} |f^p(\omega)|
\]

where \( C_p \) is a constant independent of \( j, k \) and \( f \) and \( I_{jk} = \text{Supp}\left\{ \psi_{jk} \right\} = \left[k \frac{2}{j}, k+N-1 \frac{2}{j} \right] \).

Notice that
\begin{align*}
d_{jk} & = \int \limits_{l_{j,k}} f(x) \psi_{jk}(x) dx, \\
f(x) & = \sum_{p=0}^{n-1} f^p \left( \frac{k}{2^j} \right)^p \frac{(x - \frac{k}{2^j})^p}{p!} + \frac{1}{2^j} (x - \frac{k}{2^j})^p, \omega \in \left[ \frac{k}{2^j}, x \right].
\end{align*}

**Theorem 11.** If \( \psi \) has \( p \) vanishing moments, then
\begin{enumerate}
  \item a) \( H(0) = 1 \).
  \item b) \( \frac{d^p}{d\omega^p} H(\omega) \big|_{\omega=\pi} = 0 \), \( p = 0, 1, \ldots, p - 1 \).
\end{enumerate}

**Corollary.**
\[
H(n\pi) = \begin{cases} 
1 & \text{even} \\
0 & \text{odd}
\end{cases}
\]

**Lemma 10.**
\[
\phi(2\pi n) = \delta_{0,n}, n \in \mathbb{Z}.
\]

**6. Generalized Moments**

**Probability Densities.** The functions \( \phi^2(x) \) and \( \psi^2(x) \) of orthonormal scaling function and wavelet are the probabilities densities.

**Generalized Moments.** The generalized moments of \( \phi(x) \) and \( \psi(x) \) are defined as:
\begin{align*}
\mu_{k,t} & = \int \limits_{\mathbb{R}} x^k \phi(x) \phi(x-t) dt \\
\xi_{k,t} & = \int \limits_{\mathbb{R}} x^k \psi(x) \psi(x-t) dt
\end{align*}
\( \mu_{1,0}, \xi_{1,0} \) are the first moments \([72]\) (the means) of \( \phi(x) \) and \( \psi(x) \) respectively.
\[ \varphi(x) = \sum_n h_n \sqrt{2} \varphi(2x - n), \quad \psi(x) = \sum_n g_n \sqrt{2} \varphi(2x - n), \]

**Theorem 12** ([109], see also [15]). Let \( T = 2N - 2 \). The vector \( \mu_k = \{ \mu_{k,t} \} \), \( t \leq T \), is a solution of the system

\[ \left( I - \frac{1}{2k} A \right) \mu_k = b_k, \]

where \( A_{ij} = \sum h_n h_{n+i-2j}, \quad -T \leq i, j \leq T \)

is the translation matrix (or Lawton matrix). The vector \( b_k \) has components

\[ b_{k,i} = \frac{1}{2k} \sum_i \sum_n h_i h_n \sum_{j=0}^k \binom{k}{j} n^i M_{k-j,1-n+i+2j}, \quad -T \leq i, j \leq T. \]

**Remark.** \( \mu_{1,t} = \mu_{1,-t} \).

**Note.** \( \mu_{0,t} = \delta(t) \).

**Theorem 13** [113].

\[ \mu_{1,0} = \frac{1}{2} \sum_i p_i \mu_{1,i} + \frac{1}{2} \sum_i i h_i^2, \]

where satisfy \( p_n = \sum_i \sum_n h_i h_{n+i} \), \( p_{2k} = \delta(k) \) and \( p_n = p_{-n} \).

**Theorem 14** [113]. The mean \( \xi_{1,0} = \int x \psi^2(x) \, dx \) is at the center of support of \( \psi(x) \), i.e. equal to \( \frac{2N-1}{2} \).