Chapter 1
Brief Historical Introduction

If you wish to foresee the future of mathematics our proper course is to study the history and present condition of the science.

Henri Poincaré

1.1 Fourier Series and Fourier Transforms

Historically, Joseph Fourier (1770–1830) first introduced the remarkable idea of expansion of a function in terms of trigonometric series without giving any attention to rigorous mathematical analysis. The integral formulas for the coefficients of the Fourier expansion were already known to Leonardo Euler (1707–1783) and others. In fact, Fourier developed his new idea for finding the solution of heat (or Fourier) equation in terms of Fourier series so that the Fourier series can be used as a practical tool for determining the Fourier series solution of partial differential equations under prescribed boundary conditions. Thus, the Fourier series of a function $f(x)$ defined on the interval $(-\ell, \ell)$ is given by

$$f(x) = \sum_{n=-\infty}^{\infty} c_n \exp\left(\frac{in\pi x}{\ell}\right), \tag{1.1.1}$$

where the Fourier coefficients are

$$c_n = \frac{1}{2\ell}\int_{-\ell}^{\ell} f(t) \exp\left(-\frac{in\pi x}{\ell}\right). \tag{1.1.2}$$

In order to obtain a representation for a non-periodic function defined for all real x, it seems desirable to take limit as $\ell \to \infty$ that leads to the formulation of the famous Fourier integral theorem:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega x} d\omega \int_{-\infty}^{\infty} e^{-i\omega t} f(t)\, dt. \tag{1.1.3}$$

© Springer Science+Business Media New York 2015

L. Debnath, F.A. Shah, *Wavelet Transforms and Their Applications*,
DOI 10.1007/978-0-8176-8418-1_1

Mathematically, this is a continuous version of the completeness property of Fourier series. Physically, this form (1.1.3) can be resolved into an infinite number of harmonic components with continuously varying frequency $\left(\frac{\omega}{2\pi}\right)$ and amplitude,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t} f(t)\,dt, \tag{1.1.4}$$

whereas the ordinary Fourier series represents a resolution of a given function into an infinite but discrete set of harmonic components. The most significant method of solving partial differential equations in closed form, which arose from the work of P.S. Laplace (1749–1827), was the Fourier integral. The idea is due to Fourier, A.L. Cauchy (1789–1857), and S.D. Poisson (1781–1840). It seems impossible to assign priority for this major discovery, because all three presented papers to the Academy of Sciences of Paris simultaneously. They also replaced the Fourier series representation of a solution of partial differential equations of mathematical physics by an integral representation and thereby initiated the study of Fourier integrals. At any rate, the Fourier series and Fourier integrals, and their applications were the major topics of Fourier's famous treatise entitled *Théore Analytique de la Chaleur* (The Analytic Theory of Heat) published in 1822.

In spite of the success and impact of Fourier series solutions of partial differential equations, one of the major efforts, from a mathematical point of view, was to study the problem of convergence of Fourier series. In his seminal paper of 1829, P.G.L. Dirichlet (1805–1859) proved a fundamental theorem of pointwise convergence of Fourier series for a large class of functions. His work has served as the basis for all subsequent developments of the theory of Fourier series which was profoundly a difficult subject. G.F.B. Riemann (1826–1866) studied under Dirichlet in Berlin and acquired an interest in Fourier series. In 1854, he proved necessary and sufficient conditions which would give convergence of a Fourier series of a function. Once Riemann declared that Fourier was the first who understood the nature of trigonometric series in an exact and complete manner. Later on, it was recognized that the Fourier series of a continuous function may diverge on an arbitrary set of measure zero. In 1926, A.N. Kolmogorov proved that there exists a Lebesgue integrable function whose Fourier series diverges everywhere. The fundamental question of convergence of Fourier series was resolved by L. Carleson in 1966 who proved that the Fourier series of a continuous function converges almost everywhere.

In view of the abundant development and manifold applications of the Fourier series and integrals, the fundamental problem of series expansion of an arbitrary function in terms of a given set of functions has inspired a great deal of modern mathematics.

The Fourier transform originated from the Fourier integral theorem that was stated in the Fourier treatise entitled *La Théore Analytique de la Chaleur*, and its deep significance has subsequently been recognized by mathematicians and physicists. It is generally believed that the theory of Fourier series and Fourier transforms is one of the most remarkable discoveries in mathematical sciences and has widespread applications in mathematics, physics, and engineering. Both Fourier series and Fourier transforms are related in many important ways. Many applications, including the analysis of stationary signals and real-time signal processing, make an effective use of the Fourier transform in time and frequency domains. The Fourier transform of a signal or function $f(t)$ is defined by

$$\mathscr{F}\{f(t)\} = \hat{f}(\omega) = \int_{-\infty}^{\infty} \exp(-i\omega t) f(t)\, dt = \left\langle f, e^{i\omega t} \right\rangle, \tag{1.1.5}$$

where $\hat{f}(\omega)$ is a function of frequency ω and $\langle f, e^{i\omega t}\rangle$ is the inner product in a Hilbert space. Thus, the transform of a signal decomposes it into a sine wave of different frequencies and phases, and it is often called the *Fourier spectrum.*

The remarkable success of the Fourier transform analysis is due to the fact that, under certain conditions, the signal $f(t)$ can be reconstructed by the Fourier inverse formula

$$f(t) = \mathscr{F}^{-1}\left\{\hat{f}(\omega)\right\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(i\omega t)\hat{f}(\omega)\, d\omega = \frac{1}{2\pi}\left\langle \hat{f}, e^{-i\omega t}\right\rangle. \tag{1.1.6}$$

Thus, the Fourier transform theory has been very useful for analyzing harmonic signals or signals for which there is no need for local information.

On the other hand, Fourier transform analysis has also been very useful in many other areas, including quantum mechanics, wave motion, and turbulence. In these areas, the Fourier transform $\hat{f}(k)$ of a function $f(x)$ is defined in the space and wavenumber domains, where x represents the space variable and k is the wavenumber. One of the important features is that the trigonometric kernel $\exp(-ikx)$ in the Fourier transform oscillates indefinitely, and hence, the localized information contained in the signal $f(x)$ in the x-space is widely distributed among $\hat{f}(k)$ in the Fourier transform space. Although $\hat{f}(k)$ does not lose any information of the signal $f(x)$, it spreads out in the k-space. If there are computational or observational errors involved in the signal $f(x)$, it is almost impossible to study its properties from those of $\hat{f}(k)$.

In spite of some remarkable success, Fourier transform analysis seems to be inadequate for studying above physical problems for at least two reasons. First, the Fourier transform of a signal does not contain any local information in the sense that it does not reflect the change of wavenumber with space or of frequency with time. Second, the Fourier transform method enables us to investigate problems either in time (space) domain or the frequency (wavenumber) domain, but not simultaneously in both domains. These are probably the major weaknesses of the Fourier transform analysis. It is often necessary to define a single transform of time and frequency (or space and wavenumber) that can be used to describe the energy density of a signal simultaneously in both time and frequency domains. Such a signal transform would give complete time and frequency (or space and wavenumber) information of a signal.

1.2 Gabor Transforms

In quantum mechanics, the Heisenberg uncertainty principle states that the position and momentum of a particle described by a wave function $\psi \in L^2(\mathbb{R})$ cannot be simultaneously and arbitrarily small. Motivated by this principle in 1946, Dennis Gabor (1900–1979), a Hungarian-British physicist and engineer who won the 1971 Nobel Prize in physics for his great investigation and development of holography, first recognized the great importance of localized time and frequency concentrations in signal processing. He then introduced the windowed Fourier transform to measure localized frequency components of sound waves. According to the Heisenberg uncertainty principle, the energy spread of a signal and its Fourier transform cannot be simultaneously and arbitrarily small. Gabor first identified a signal with a family of waveforms which are well concentrated in time and in frequency. He called these elementary waveforms as the *time–frequency atoms* that have a minimal spread in a time–frequency plane.

In fact, Gabor formulated a fundamental method for decomposition of signals in terms of elementary signals (or atomic waveforms). His pioneering approach has now become one of the standard models for time–frequency signal analysis.

In order to incorporate both time and frequency localization properties in one single transform function, Gabor first introduced the *windowed Fourier transform* (or the Gabor transform) by using a Gaussian distribution function as a window function. His major idea was to use a time-localization window function $g_a(t-b)$ for extracting local information from the Fourier transform of a signal, where the parameter a measures the width of the window, and the parameter b is used to translate the window in order to cover the whole time domain. The idea is to use this window function in order to localize the Fourier transform, then shift the window to another position, and so on. This remarkable property of the Gabor transform provides the local aspect of the Fourier transform with time resolution equal to the size of the window. In fact, Gabor (1946) used $g_{t,\omega}(\tau) = \bar{g}(\tau - t)\exp(i\omega\tau)$ as the window function by translating and modulating a function g, where $g(\tau) = \pi^{-\frac{1}{4}}\exp\left(-2^{-1}\tau^2\right)$, which is the so-called canonical coherent states in quantum physics. The *Gabor transform* (*windowed Fourier transform*) of f with respect to g, denoted by $\tilde{f}_g(t,\omega)$, is defined as

$$\mathscr{G}[f](t,\omega) = \tilde{f}_g(t,\omega) = \int_{-\infty}^{\infty} f(\tau)g(\tau - t)\,e^{-i\omega\tau}d\tau = \langle f, \bar{g}_{t,\omega}\rangle, \tag{1.2.1}$$

where $f, g \in L^2(\mathbb{R})$ with the inner product $\langle f, g\rangle$. In practical applications, f and g represent signals with finite energy. In quantum mechanics, $\tilde{f}_g(t,\omega)$ is referred to as the *canonical coherent state representation* of f. The term "coherent states" was first used by Glauber (1964) in quantum optics. The inversion formula for the Gabor transform is given by

$$\mathscr{G}^{-1}\left[\tilde{f}_g(t,\omega)\right] = f(\tau) = \frac{1}{2\pi}\frac{1}{\|g\|^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{f}_g(t,\omega)\,\bar{g}(\tau-t)\,e^{i\omega t}\,dt\,d\omega. \tag{1.2.2}$$

In terms of the sampling points defined by $t = mt_0$ and $\omega = n\omega_0$, where m and n are integers and ω_0 and t_0 are positive quantities, the discrete Gabor functions are defined by $g_{m,n}(t) = g(t - mt_0)\exp(-in\omega_0 t)$. These functions are called the *Weyl–Heisenberg coherent states*, which arise from translations and modulations of the Gabor window function. From a physical point of view, these coherent states are of special interest. They have led to several important applications in quantum mechanics. Subsequently, various other functions have been used as window functions instead of the Gaussian function that was originally introduced by Gabor. The discrete Gabor transform is defined by

$$\tilde{f}(m,n) = \int_{-\infty}^{\infty} f(t)\,\bar{g}_{m,n}(t)\,dt = \langle f, g_{m,n}\rangle. \tag{1.2.3}$$

The double series $\sum_{m,n=-\infty}^{\infty} \tilde{f}(m,n)\,g_{m,n}(t)$ is called the *Gabor series* of $f(t)$.

In many applications, it is more convenient, at least from a numerical point of view, to deal with discrete transforms rather than continuous ones. The discrete Gabor transform is defined by

$$\tilde{f}(mt_0, n\omega_0) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(\tau)\,g_{m,n}(\tau)\,d\tau = \frac{1}{\sqrt{2\pi}}\langle f, \bar{g}_{m,n}\rangle. \tag{1.2.4}$$

If the functions $\{g_{m,n}(t)\}$ form an orthonormal basis or, more generally, if they form a frame on $L^2(\mathbb{R})$, then $f \in L^2(\mathbb{R})$ can be reconstructed by the formula

$$f(t) = \sum_{m,n=-\infty}^{\infty} \langle f, g_{m,n}\rangle\, g^*_{m,n}(t), \tag{1.2.5}$$

where $\{g^*_{m,n}(t)\}$ is the dual frame of $\{g_{m,n}(t)\}$. The discrete Gabor transform deals with a discrete set of coefficients which allows efficient numerical computation of those coefficients. However, Malvar (1990a,b) recognized some serious algorithmic difficulties in the Gabor wavelet analysis. He resolved these difficulties by introducing new wavelets which are now known as the *Malvar wavelets* and fall within the general framework of the window Fourier analysis. From an algorithmic point of view, the Malvar wavelets are much more effective and superior to Gabor wavelets and other wavelets.

1.3 The Wigner–Ville Distribution and Time–Frequency Signal Analysis

In a remarkable paper, Wigner (1932), the 1963 Nobel Prize Winner in Physics, first introduced a new function $W_\psi(x, p)$ of two independent variables from the wave function ψ in the context of quantum mechanics defined by

$$W_\psi(x, p) = \frac{1}{h}\int_{-\infty}^{\infty} \psi\left(x + \frac{1}{2}t\right)\overline{\psi}\left(x - \frac{1}{2}t\right)\exp\left(\frac{ipt}{\hbar}\right)dt, \tag{1.3.1}$$

where ψ satisfies the one-dimensional Schrödinger equation, the variables x and p represent the quantum-mechanical position and momentum respectively, and $h = 2\pi\hbar$ is the Planck constant. The Wigner function $W_\psi(x, p)$ has many remarkable properties which include the space and momentum marginal integrals

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} W_\psi(x, p)\,dp = \left|\psi(x)\right|^2, \quad \frac{1}{2\pi}\int_{-\infty}^{\infty} W_\psi(x, p)\,dx = \left|\hat{\psi}(p)\right|^2. \tag{1.3.2a,b}$$

These integrals represent the usual position and momentum energy densities. Moreover, the integral of the Wigner function over the whole (x, p) space is

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W_\psi(x, p)\,dx\,dp = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\hat{\psi}(p)\right|^2 dp = \int_{-\infty}^{\infty}\left|\psi(x)\right|^2 dx. \tag{1.3.3}$$

This can be interpreted as the total energy over the whole position-momentum plane (x, p).

As is well known, the Fourier transform analysis is a very effective tool for studying stationary (time-independent) signals (or waveforms). However, signals (or waveforms) are, in general, nonstationary. Such signals or waveforms cannot be analyzed completely by the Fourier analysis. Therefore, a complete analysis of non-stationary signals (or waveforms) requires both time–frequency (or space-wavenumber) representations of signals. In 1948, Ville proposed the Wigner distribution of a function or signal $f(t)$ in the form

$$W_f(t, \omega) = \int_{-\infty}^{\infty} f\left(t + \frac{\tau}{2}\right)\bar{f}\left(t - \frac{\tau}{2}\right)e^{-i\omega\tau}d\tau, \tag{1.3.4}$$

for analysis of the time–frequency structures of nonstationary signals, where $\bar{f}(z)$ is the complex conjugate of $f(z)$. Subsequently, this time–frequency representation (1.3.4) of a signal f is known as the *Wigner–Ville distribution* (WVD) which is one of the fundamental methods that have been developed over the years for the time–frequency signal analysis. An extensive study of this distribution was made by Claasen and Mecklenbräuker (1980) in the context of the time–frequency signal analysis. Besides other linear time–frequency representations, such as the short-time

Fourier transform or the Gabor transform, and the WVD plays a central role in the field of bilinear/quadratic time–frequency representations. In view of its remarkable mathematical structures and properties, the WVD is now well recognized as an effective method for the time–frequency (or space wavenumber) analysis of non-stationary signals (or waveforms), and nonstationary random processes. In recent years, this distribution has served as a useful analysis tool in many fields as diverse as quantum mechanics, optics, acoustics, communications, biomedical engineering, signal processing, and image processing. It has also been used as a method for analyzing seismic data, and the phase distortion involved in a wide variety of audio engineering problems. In addition, it has been suggested as a method for investigating many important topics including instantaneous frequency estimation, spectral analysis of non-stationary random signals, detection and classification of signals, algorithms for computer implementation, speech signals, and pattern recognition.

In sonar and radar systems, a real signal is transmitted and its echo is processed in order to find out the position and velocity of a target. In many situations, the received signal is different from the original one only by a time translation and the Doppler frequency shift. In the context of the mathematical analysis of radar information, Woodward (1953) reformulated the theory of the WVD. He introduced a new function $A_f(t,\omega)$ of two independent variables t, ω from a radar signal f in the form

$$A_f(t,\omega) = \int_{-\infty}^{\infty} f\left(\tau + \frac{t}{2}\right) \bar{f}\left(\tau - \frac{t}{2}\right) e^{-i\omega\tau} d\tau. \tag{1.3.5}$$

This function is now known as the *Woodward ambiguity function* and plays a central role in radar signal analysis and radar design. The ambiguity function has been widely used for describing the correlation between a radar signal and its Doppler-shifted and time-translated version. It was also shown that the ambiguity function exhibits the measurement between ambiguity and target resolution, and for this reason it is also known as the *radar ambiguity function*. In analogy with the Heisenberg uncertainty principle in quantum mechanics, Woodward also formulated a *radar uncertainty principle*, which says that the range and velocity (range rate) cannot be measured exactly and simultaneously. With the activity surrounding the radar uncertainty principle, the representation theory of the Heisenberg group and ambiguity functions as special functions on the Heisenberg group led to a series of many important results. Subsequently, considerable attention has been given to the study of radar ambiguity functions in harmonic analysis and group theory by several authors, including Wilcox (1960), Schempp (1984), and Auslander and Tolimieri (1985).

From theoretical and application points of view, the WVD plays a central role and has several important and remarkable structures and properties. First, it provides a high-resolution representation in time and in frequency for some nonstationary signals. Second, it has the special property of satisfying the time and frequency marginals in terms of the instantaneous power in time and energy spectrum in frequency. Third, the first conditional moment of frequency at a given time is the derivative of the phase of the signal at that time. The derivative of the phase divided by 2π gives the *instantaneous frequency* which is uniquely related to the signal. Moreover, the second conditional moment of frequency of a signal does not have any physical interpretation. In spite of these remarkable features, its energy distribution is *not* nonnegative and it often possesses severe cross-terms, or interference terms between different time–frequency regions, leading to undesirable properties.

In order to overcome some of the inherent weaknesses of the WVD, there has been considerable recent interest in more general time–frequency distributions as a mathematical method for time–frequency signal analysis. Often, the WVD has been modified by smoothing in one or two dimensions, or by other signal processing. In 1966, Cohen introduced a general class of bilinear shift-invariant, quadratic time–frequency distributions in the form

$$C_f(t,\nu) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left[-2\pi i(\nu\tau + st - rs)\right] \times g(s,\tau) f\left(r+\frac{\tau}{2}\right)\bar{f}\left(r-\frac{\tau}{2}\right) e^{-i\omega\tau}\, d\tau\, dr\, ds, \tag{1.3.6}$$

where the given kernel $g(s,\tau)$ generates different distributions which include windowed Wigner–Ville, Choi–Williams, spectrogram, Rihaczek, Born–Jordan, and Page distributions. In modern time–frequency signal analysis, several alternative forms of the Cohen distribution seem to be convenient and useful. A function u is introduced in terms of the given kernel $g(s,\tau)$ by

$$u(r,\tau) = \int_{-\infty}^{\infty} g(s,\tau)\exp(2\pi i s r)\, ds \tag{1.3.7}$$

so that the Cohen distribution takes the general form

$$C_f(t,\nu) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(r-t,\tau) f\left(r+\frac{\tau}{2}\right)\bar{f}\left(r-\frac{\tau}{2}\right)\exp(-2\pi i\nu\tau)\, d\tau\, dr. \tag{1.3.8}$$

The general Cohen distribution can also be written in terms of an ambiguity function as

$$C_f(t,\nu) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(s,\tau)\exp\left[-2\pi i(st+\nu\tau)\right] ds\, d\tau, \tag{1.3.9}$$

where $A(s,\tau)$ is the general ambiguity function of f and g defined by

$$A(s,\tau) = g(s,\tau)\int_{-\infty}^{\infty} f\left(r+\frac{\tau}{2}\right)\bar{f}\left(r-\frac{\tau}{2}\right)\exp(2\pi i r s)\, dr. \tag{1.3.10}$$

As a natural generalization of the WVD, another family of bilinear time–frequency representations was introduced by Rihaczek in 1968. This is called the *generalized Wigner–Ville* (GWV) distribution or more appropriately, the *Wigner–Ville–Rihaczek* (WVR) distribution which is defined for two signals f and g by

$$R_{f,g}^{\alpha}(t,\omega) = \int_{-\infty}^{\infty} f\left(t + \left(\frac{1}{2} - \alpha\right)\tau\right) \bar{g}\left(t - \left(\frac{1}{2} + \alpha\right)\tau\right) e^{-i\omega\tau} d\tau, \tag{1.3.11}$$

where α is a real constant parameter. In particular, when $\alpha = 0$, (1.3.11) reduces to the WVD, and when $\alpha = 2^{-1}$, (1.3.11) represents the *Wigner–Rihaczek* distribution in the form

$$R_{f,g}^{\frac{1}{2}}(t,\omega) = f(t) \int_{-\infty}^{\infty} \bar{g}(t-\tau) e^{-i\omega\tau} d\tau = f(t)\, e^{-i\omega t}\, \hat{\bar{g}}(\omega). \tag{1.3.12}$$

The main feature of these distributions is their time- and frequency-shift invariance. However, for some problems where the scaling of signals is important, it is necessary to consider distributions which are invariant to translations and compressions of time, that is, $t \to at + b$ (affine transformations). Bertrand and Bertrand (1992) obtained another general class of distributions which are called *affine time–frequency distributions* because they are invariant to affine transformations. Furthermore, extended forms of the various affine distributions are also introduced to obtain representations of complex signals on the whole time–frequency plane. The use of the real signal in these forms shows the effect of producing symmetry of the result obtained with the analytic signal. In any case, the construction based on the affine group, which is basic in signal analysis, ensures that no spurious interference will ever occur between positive and negative frequencies. Special attention has also been given to the computational aspects of broadband functionals containing stretched forms of the signal such as affine distributions, wavelet coefficients, and broadband ambiguity functions. Different methods based on group theory have also been developed to derive explicit representations of joint time–frequency distributions adapted to the analysis of wideband signals.

Although signal analysis originated more than 50 years ago, there has been major development of the time–frequency distributions approach in the basic idea of the method to develop a joint function of time and frequency, known as a time–frequency distribution, that can describe the energy density of a signal simultaneously in both time and frequency domains. In principle, the joint time–frequency distributions characterize phenomena in the two-dimensional time–frequency plane. Basically, there are two kinds of time–frequency representations. One is the quadratic method describing the time–frequency distributions, and the other is the linear approach including the Gabor transform and the wavelet transform. Thus, the field of time–frequency analysis has evolved into a widely recognized applied discipline of signal processing over the last two decades. Based on studies of its mathematical structures and properties by many authors including De Bruijn (1967, 1973), Claasen and Mecklenbräuker (1980), Boashash (1992), Mecklenbräuker and Hlawatsch (1997), the WVD and its various generalizations with applications were brought to the attention of larger mathematical, scientific, and engineering communities. By any assessment, the WVD has served as the fundamental basis for all subsequent classical and modern developments of time–frequency signal analysis and signal processing.

1.4 Wavelet Transforms

Historically, the concept of “ondelettes” or “wavelets” started to appear more frequently only in the early 1980s. This new concept can be viewed as a synthesis of various ideas originating from different disciplines including mathematics (Calderón–Zygmund operators and Littlewood–Paley theory), physics (the coherent states formalism in quantum mechanics and the renormalization group), and engineering (quadratic mirror filters (QMF), sideband coding in signal processing, and pyramidal algorithms in image processing). In 1982, Jean Morlet, a French geophysical engineer, discovered the idea of the wavelet transform, providing a new mathematical tool for seismic wave analysis. In Morlet’s analysis, signals consist of different features in time and frequency, but their high-frequency components would have a shorter time duration than their low-frequency components. In order to achieve good time resolution for the high-frequency transients and good frequency resolution for the low-frequency components, Morlet et al. (1982a,b) first introduced the idea of wavelets as a family of functions constructed from translations and dilations of a single function called the “mother wavelet” $\psi(t)$. They are defined by

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\,\psi\left(\frac{t-b}{a}\right), \quad a, b \in \mathbb{R},\ a \neq 0, \tag{1.4.1}$$

where a is called a scaling parameter which measures the degree of compression or scale, and b a translation parameter which determines the time location of the wavelet. If $|a| < 1$, the wavelet (1.4.1) is the compressed version (smaller support in time-domain) of the mother wavelet and corresponds mainly to higher frequencies. On the other hand, when $|a| > 1$, $\psi_{a,b}(t)$ has a larger time-width than $\psi(t)$ and corresponds to lower frequencies. Thus, wavelets have time-widths adapted to their frequencies. This is the main reason for the success of the Morlet wavelets in signal processing and time–frequency signal analysis. It may be noted that the resolution of wavelets at different scales varies in the time and frequency domains as governed by the Heisenberg uncertainty principle. At large scale, the solution is coarse in the time domain and fine in the frequency domain. As the scale a decreases, the resolution in the time domain becomes finer while that in the frequency domain becomes coarser.

Morlet first developed a new time–frequency signal analysis using what he called “wavelets of constant shape” in order to contrast them with the analyzing functions in the short-time Fourier transform which do not have a constant shape. It was Alex Grossmann, a French theoretical physicist, who quickly recognized the importance of the MorIet wavelet transforms which are somewhat similar to the formalism for coherent states in quantum mechanics, and developed an exact inversion formula for this wavelet transform. Unlike the Weyl–Heisenberg coherent states, these coherent states arise from translations and dilations of a single function. They are often called *affine coherent states* because they are associated with an affine group (or “$ax + b$”

group). From a group-theoretic point of view, the wavelets $\psi_{a,b}(x)$ are in fact the result of the action of the operators $U(a,b)$ on the function ψ so that

$$\left[U(a,b)\psi\right](x) = \frac{1}{\sqrt{|a|}}\,\psi\left(\frac{x-b}{a}\right). \tag{1.4.2}$$

These operators are all unitary on the Hilbert space $L^2(\mathbb{R})$ and constitute a representation of the "$ax + b$" group:

$$U(a,b)\,U(c,d) = U(ac, b + ad). \tag{1.4.3}$$

This group representation is *irreducible*, that is, for any nonzero $f \in L^2(\mathbb{R})$, there exists no nontrivial g orthogonal to all the $U(a,b)f$. In other words, $U(a,b)f$ span the entire space. The coherent states for the affine $(ax + b)$-group, which are now known as *wavelets*, were first formulated by Aslaksen and Klauder (1968, 1969) in the context of more general representations of groups. The success of Morlet's numerical algorithms prompted Grossmann to make a more extensive study of the Morlet wavelet transform which led to the recognition that wavelets $\psi_{a,b}(t)$ correspond to a square integrable representation of the affine group. Grossmann was concerned with the wavelet transform of $f \in L^2(\mathbb{R})$ defined by

$$\mathscr{W}_\psi\,[f]\,(a,b) = \left\langle f, \psi_{a,b}\right\rangle = \frac{1}{\sqrt{|a|}}\int_{-\infty}^{\infty} f(t)\,\overline{\psi\left(\frac{t-b}{a}\right)}dt, \tag{1.4.4}$$

where $\psi_{a,b}(t)$ plays the same role as the kernel $\exp(i\omega t)$ in the Fourier transform. Like the Fourier transformation, the continuous wavelet transformation $\mathscr{W}_\psi$ is linear. However, unlike the Fourier transform, the continuous wavelet transform is not a single transform, but any transform obtained in this way. The inverse wavelet transform can be defined so that f can be reconstructed by means of the formula

$$f(t) = C_\psi^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathscr{W}_\psi\,[f]\,(a,b)\psi_{a,b}(t)\left(a^{-2}da\right)db, \tag{1.4.5}$$

provided C_ψ satisfies the so-called admissibility condition

$$C_\psi = 2\pi\int_{-\infty}^{\infty}\frac{\left|\hat{\psi}(\omega)\right|^2}{|\omega|}\,d\omega < \infty, \tag{1.4.6}$$

where $\hat{\psi}(\omega)$ is the Fourier transform of the mother wavelet $\psi(t)$. Grossmann's ingenious work also revealed that certain algorithms that decompose a signal on the whole family of scales, can be utilized as an efficient tool for multiscale analysis. In practical applications involving fast numerical algorithms, the continuous wavelet can be computed at discrete grid points. To do this, a general wavelet ψ can be defined by replacing a with $a_0^m (a_0 \neq 0, 1), b$ with $n b_0 a_0^m (b_0 \neq 0)$, where m and n are integers, and making

$$\psi_{m,n}(t) = a_0^{-m/2} \psi\big(a_0^{-m} t - n b_0\big). \tag{1.4.7}$$

The discrete wavelet transform (DWT) of f is defined as the doubly indexed sequence

$$\tilde{f}(m,n) = \mathscr{W}[f](m,n) = \langle f, \psi_{m,n} \rangle = \int_{-\infty}^{\infty} f(t)\, \overline{\psi}_{m,n}(t)\, dt, \tag{1.4.8}$$

where $\psi_{m,n}(t)$ is given by (1.4.7). The double series

$$\sum_{m,n=-\infty}^{\infty} \tilde{f}(m,n)\, \psi_{m,n}(t), \tag{1.4.9}$$

is called the *wavelet series* of f, and the functions $\{\psi_{m,n}(t)\}$ are called the *discrete wavelets*, or simply *wavelets*. However, there is no guarantee that the original function f can be reconstructed from its discrete wavelet coefficients in general. The reconstruction of f is still possible if the discrete lattice has a very fine mesh. For very coarse meshes, the coefficients may not contain sufficient information for determination of f from these coefficients. However, for certain values of the lattice parameter (m, n), a numerically stable reconstruction formula can be obtained. This leads to the concept of a "frame" rather than bases. The notion of the frame was introduced by Duffin and Schaeffer (1952) for the study of a class of nonharmonic Fourier series to which Paley and Wiener made fundamental contributions. They discussed related problems of nonuniform sampling for band-limited functions.

In general, the function f belonging to the Hilbert space, $L^2(\mathbb{R})$ (see Debnath and Mikusinski 1999), can be completely determined by its DWT (wavelet coefficients) if the wavelets form a complete system in $L^2(\mathbb{R})$. In other words, if the wavelets form an orthonormal basis or a frame of $L^2(\mathbb{R})$, then they are complete. And f can be reconstructed from its DWT $\left\{\tilde{f}(m,n) = \langle f, \psi_{m,n} \rangle\right\}$ by means of the formula

$$f(x) = \sum_{m,n=-\infty}^{\infty} \langle f, \psi_{m,n} \rangle\, \psi_{m,n}(x), \tag{1.4.10}$$

provided the wavelets form an orthonormal basis.

On the other hand, the function f can be determined by the formula

$$f(x) = \sum_{m,n=-\infty}^{\infty} \langle f, \psi_{m,n} \rangle\, \tilde{\psi}_{m,n}(x), \tag{1.4.11}$$

provided the wavelets form a frame and $\{\tilde{\psi}_{m,n}(x)\}$ is the dual frame.

For some very special choices of ψ and a_0, b_0, the $\psi_{m,n}$ constitute an orthonormal basis for $L^2(\mathbb{R})$. In fact, if $a_0 = 2$ and $b_0 = 1$, then there exists a function ψ with good time–frequency localization properties such that

$$\psi_{m,n}(x) = 2^{-m/2}\psi\big(2^{-m}x - n\big) \tag{1.4.12}$$

form an orthonormal basis for $L^2(\mathbb{R})$. These $\{\psi_{m,n}(x)\}$ are known as the *Littlewood–Paley wavelets*. This gives the following representation of f

$$f(x) = \sum_{m,n=-\infty}^{\infty} \langle f, \psi_{m,n}\rangle \psi_{m,n}(x), \tag{1.4.13}$$

which has a good space-frequency localization. The classic example of a wavelet ψ for which the $\psi_{m,n}$ defined by (1.4.12) constitute an orthonormal basis for $L^2(\mathbb{R})$ is the Haar wavelet

$$\psi(x) = \begin{cases} 1, & 0 \le x < \frac{1}{2} \\ -1, & \frac{1}{2} \le x < 1 \\ 0, & \text{otherwise} \end{cases} \tag{1.4.14}$$

Historically, the first orthonormal wavelet basis is the Haar basis, which was discovered long before the wavelet was introduced. It may be observed that the Haar wavelet ψ does not have good time–frequency localization and that its Fourier transform $\hat{\psi}(k)$ decays like $|k|^{-1}$ as $k \to \infty$. The joint venture of Morlet and Grossmann led to a detailed mathematical study of the wavelet transforms and their applications. It became clear from their work that, analogous to the Fourier expansion of functions, the wavelet transform analysis provides a new method for decomposing a function (or a signal).

In 1985, Yves Meyer, a French pure mathematician, recognized the deep connection between the Calderón formula in harmonic analysis and the new algorithm discovered by Grossmann and Morlet (1984). He also constructed an orthonormal basis, for the Hilbert space $L^2(\mathbb{R})$, of wavelets $\psi_{m,n}$ defined by (1.4.12) based on the mother wavelet ψ with compact support and C^∞ Fourier transform $\hat{\psi}$. This basis turned out to be an unconditional basis for all L^p spaces $(1 < p < \infty)$, Sobolev spaces, and other spaces. Furthermore, in a Hilbert space, a normalized basis turns out to be an unconditional basis if and only if it is also a frame. Such a basis is called the *Riesz basis*. However, if $\{\psi_n\}$ is an orthonormal basis, then $e_n = \big(1 + n^2\big)^{-\frac{1}{2}}(n\psi_1 + \psi_2)$ is an example of a basis of normalized vectors that is not a Riesz basis. Using the knowledge of the Calderón–Zygmund operators and the Littlewood–Paley theory, in 1985–1986 Meyer (1990) successfully gave a mathematical foundation of the wavelet theory. The Meyer basis has become a more powerful tool than the Haar basis.

Even though the mother wavelet in the Meyer basis decays faster than any inverse polynomials, the constants involved are very large so that it is not very well localized. Lemarié and Meyer (1986) extended the Meyer orthonormal basis to more than one dimension. One of the new orthonormal wavelet bases for $L^2(\mathbb{R})$ with localization properties in both time and frequency was first constructed by Strömberg in 1982. His wavelets are in C^n, where n is arbitrary but finite and decays exponentially. He also proved that the orthonormal wavelet basis defined by (1.4.12) is, in fact, an unconditional basis for the Hardy space $\mathscr{H}^1(\mathbb{R})$ which consists of real-valued functions $u(x)$ if and only if $u(x)$ and its Hilbert transform $\hat{u}(\kappa)$ belong to $L^1(\mathbb{R})$. In fact, $\mathscr{H}^1(\mathbb{R})$ is the real version of the holomorphic Hardy space $\mathscr{H}^1(\mathbb{R})$ whose elements are $u(x) + i v(x)$, where $u(x)$ and $v(x)$ are real-valued functions. A function $f(z)$, where $z = x + iy$, belongs to the Hardy space $H^p(\mathbb{R}), 0 \le p \le \infty$, if it is holomorphic in the upper half ($y > 0$) of the complex plane and if

$$\|f\|_p = \sup_{y>0} \left[\int_{-\infty}^{\infty} |f(z)|^p dx\right]^{\frac{1}{p}} < \infty. \tag{1.4.15}$$

If this condition is satisfied, the upper bound, taken over $y > 0$, is also the limit as $y \to 0$. Moreover, $f(z)$ converges to a function $f(x)$ as $y \to 0+$, where convergence is in the sense of the L^p-norm. The space $H^p(\mathbb{R})$ can thus be identified with a closed subspace of $L^p(\mathbb{R})$. The Hardy space $H^2(\mathbb{R})$ plays a major role in signal processing. The real part of an analytic signal $F(t) = f(t) + i g(t),\ t \in \mathbb{R}$, represents a real signal $f(t)$ with finite energy given by

$$\|f\| = \left[\int_{-\infty}^{\infty} |f(t)|^2 dt\right]^{\frac{1}{2}}. \tag{1.4.16}$$

If F has finite energy, then $F \in H^2(\mathbb{R})$.

The *fractional Fourier transform* (FRFT) is a generalization of the ordinary Fourier transform with an order parameter α and is identical to the ordinary Fourier transform when this order α is equal to $\alpha = \pi/2$. However, this transform has one major drawback due to using global kernel i.e., it only provides such FRFT spectral content with no indication about the time localization of the FRFT spectral components. Therefore, the analysis of non-stationary signals whose FRFT spectral characteristics change with time requires joint signal representations in both time and FRFT domains, rather than just a FRFT domain representation. The first modification to the FRFT to allow analysis of aforementioned non-stationary signals came as the *short-time FRFT* (STFRFT). But the short coming of this transform is that its time and fractional-domain resolutions cannot simultaneously be arbitrarily high. As a generalization of the wavelet transform, Mendlovic et al. (1997) first introduced the *fractional wavelet transform* (FRWT) as a way to deal with optical signals. The FRWT with an order α of any function $f(t) \in L^2(\mathbb{R})$ is defined as

$$\mathscr{W}_\psi^\alpha[f](a,b) = \int_{\mathbb{R}} f(t)\,\overline{\psi_{a,b,\alpha}(t)}\,dt \tag{1.4.17}$$

where

$$\psi_{a,b,\alpha}(t) = \frac{1}{\sqrt{a}}\,\psi\left(\frac{t-b}{a}\right) e^{-j\,\frac{t^2-b^2}{2}\,\cot\alpha}, \quad a \in \mathbb{R}^+, b \in \mathbb{R}. \tag{1.4.18}$$

Note that when $\alpha = \pi/2$, the FRWT reduces to the classical wavelet transform. The idea behind this transform is deriving the fractional spectrum of the signal by using the FRFT and performing the wavelet transform of the fractional spectrum. Besides being a generalization of the wavelet transform, the FRWT can be interpreted as a rotation of the time–frequency plane and has been proved to relate to other time-varying signal analysis tools, which make it as a unified time–frequency transform. In recent years, this transform has been paid a considerable amount of attention, resulting in many applications in the areas of optics, quantum mechanics, pattern recognition and signal processing.

The *inverse FRWT* can be defined so that f can be reconstructed by the formula

$$f(t) = \frac{1}{2\pi\,C_\psi}\int_{\mathbb{R}} \mathscr{W}_\psi^\alpha[f](a,b)\,\psi_{a,b,\alpha}(t)\,\frac{da\,db}{a^2} \tag{1.4.19}$$

provided C_ψ satisfies the admissibility condition (1.4.6).

A comprehensive overview of FRFT and FRWTs can also be found in Almeida (1994), Mendlovic et al. (1997), Ozaktas et al. (2000) and Sejdić et al. (2011).

Although the DWT has established an impressive reputation as a tool for mathematical analysis and signal processing, it suffers from three major disadvantages: *shift sensitivity, poor directionality*, and *lack of phase information*. These disadvantages severely restrict its scope for certain signal and image processing applications, for example, edge detection, image registration/segmentation, motion estimation. Significant progress in the development of directional wavelets has been made in recent years. There are certain applications for which the optimal representation can be achieved through more redundant extensions of standard DWT such as *wavelet packet transform* (WPT) and *stationary wavelet transform* (SWT). All these forms of DWTs result in real valued transform coefficients with two or more limitations. There is an alternate way of reducing these limitations with a limited redundant representation in complex domain. In fact, the initial motivation behind the earlier development of *complex-valued wavelet transform* (CWT) was the third limitation that is the "absence of phase information". Complex wavelets transforms use complex-valued filtering (analytic filter) that decomposes the real/complex signals into real and imaginary parts in transform domain (see Lawton 1993). The real and imaginary coefficients are used to compute amplitude and phase information, just the type of information needed to accurately describe the energy localization of wavelet functions. As such complex wavelet transform is one way to improve directional selectivity and only requires $O(N)$ computational cost. However, the complex wavelet transform has not been widely used in the past, since it is difficult to design complex wavelets with perfect reconstruction properties and good filter characteristics (see Fernandes 2002; Gao et al. 2002; Neumann and Steidl 2005).

Another popular technique is the *dual-tree complex wavelet transform* (DTCWT) proposed by Kingsbury (1999, 2001), which added perfect reconstruction to the other attractive properties of complex wavelets, including approximate shift invariance, six directional selectivities, limited redundancy and efficient $O(N)$ computation. On the other hand, Selesnick (2001) proposed *dual-density DTCWT* (DDTCWT), an alternative filter design methods for DTCWT almost equivalent to Kingsbury's transform such that in the limit, the scaling and wavelet functions form Hilbert transform pairs. This type of transform is designed with simple methods to obtain filter coefficients. Although the dual-tree DWT based complex wavelet transform reduce all three disadvantages of standard DWT, the redundancy (though limited) of the transform is a major drawback for applications like image compression and image restoration. To overcome this disadvantage, Spaendonck et al. (2000) proposed a *non-redundant complex wavelet transform* (NRCWT) based on projections with no-redundancy for both real and complex-valued signals.

The last two decades have seen tremendous activity in the development of new mathematical and computational tools based on multiscale ideas such as *steerable wavelets, wedgelets, beamlets, bandlets, ridgelets, curvelets, contourlets, surfacelets, shearlets, and platelets*. These geometric wavelets or directional wavelets are uniformly called *X-lets*. The main advantage of these new wavelets lies in the fact that they possess all the advantages of classical wavelets, that is space localization and scalability, but additionally the geometrical wavelet transforms have strong directional character. They allow to catch changes of a signal in different directions. So we have one more parameter next to space and scalability, that is direction.

The steerable wavelets can be seen as early directional wavelets introduced by Freeman and Adelson (1991). The steerable wavelets were built based on directional derivative operators (the second derivative of a Gaussian). They provide translation invariant and rotation invariant representations of the position and the orientation of considered image structures. Since, wavelets have been very successful in applications such as denoising and compact approximations of images containing zero dimensional or point singularities. Wavelets do not isolate the smoothness along edges that occurs in images, and they are thus more appropriate for the reconstruction of sharp point singularities than lines or edges. So there was a need to create the theory, which could remedy the problem of representation of edges present in images in efficient way. Such theory was first described by Donoho (2000). He developed an overcomplete collection of atoms which are dyadically organized indicator functions with a variety of locations, scales, and orientations, and named them *wedgelets*. They were used to represent the class of smooth images with discontinuities along smooth curves in a very efficient and sparse way.

There were several other research groups working with the same goal, namely a better analysis and an optimal representation of directional features of signals in higher dimensions. To overcome the weakness of wavelets in higher dimensions, Candés and Donoho (1999a,b) pioneered a new system of representations named *ridgelets* that deal effectively with line singularities in two dimensions. Ridgelets are different from wavelets in a sense that ridgelets exhibit very high directional sensitivity and are highly anisotropic.

For $a, b \in \mathbb{R}, a > 0$ and each $\theta \in [0, 2\pi)$, the bivariate ridgelet $\psi_{a,b,\theta} : \mathbb{R}^2 \to \mathbb{R}$ is defined as

$$\psi_{a,b,\theta}(x) = \frac{1}{\sqrt{a}}\, \psi\left(\frac{x\cos\theta + y\sin\theta - b}{a}\right), \tag{1.4.20}$$

where ψ is the smooth function with sufficient decay and satisfying the admissibility condition (1.4.6). Therefore, to a certain degree, the ridgelet is a novel version of wavelet function with additional orientation information. Given an integrable bivariate function $f(x, y)$, the *continuous ridgelet transform* (CRT) of $f(x, y)$ is defined as

$$R(a, b, \theta) = \int \psi_{a,b,\theta} f(x, y)\, dx\, dy. \tag{1.4.21}$$

The ridgelet transform is also called as *anisotropic geometric wavelet transform*. The ridgelet transform can be represented in term of the *Radon transform*, which is defined as

$$RA(t, \theta) = \int f(x, y)\, \delta\, (x\cos\theta + y\sin\theta - t)\, dx\, dy, \quad t \in \mathbb{R}, \theta \in [0, \pi) \tag{1.4.22}$$

where δ is the Dirac distribution. So the ridgelet transform is precisely the application of one-dimensional wavelet transform to the slices of the Radon transform where the angular variable θ is constant and t is varying. Therefore, the basic idea of the ridgelet transform is to map a line singularity into a point singularity using the Radon transform. Then, the wavelet transform can be used to effectively handle the point singularity in the Radon domain. Thus, the ridgelet transform allows the representation of edges and other singularities along lines in a more efficient way, in terms of compactness of the representation, than the traditional transformations for a given accuracy of reconstruction.

Unfortunately, the ridgelet transform is only applicable to objects with global straight line singularities, which are rarely observed in real applications. For example, in image processing, edges are typically curved rather than straight and ridgelets alone cannot yield efficient representations. To overcome inherent limitations of this transform, Candés and Donoho (2003a) developed a new multiscale transform called *curvelet transform*, which was designed to represent edges and other singularities along curves much more efficiently than traditional transforms. The basic idea is here to partitioning the curves into collection of ridge fragments and then handle each fragment using the ridgelet transform. The curvelets are defined not only at various scales and locations but also at various orientations. Also, their supports are highly anisotropic and become increasingly elongated at finer scales. Due to those two key features, namely directionality and anisotropy, curvelets are essentially as good as an adaptive representation system from the point of view of the ability to sparsely approximate images with edges. Later on, a considerably simpler second generation curvelet transform based on a frequency partition technique was proposed by the same authors (Candés and Donoho 2003a,b). The second generation curvelet transform has been shown to be a very efficient tool for many different applications in image processing, seismic data exploration, fluid mechanics, and solving partial differential equations. Recently, a variant of the second generation curvelet transform was proposed by Demanet and Ying (2007a,b) to handle image boundaries by mirror extension.

Do and Vetterli (2003, 2005) proposed a *contourlet* transform, which provides a flexible multiresolution, local and directional expansion for images. They designed it to satisfy the anisotropy scaling relation for curves, and thus offers a fast and structured curvelet-like decomposition sampled signals. Therefore, the key difference between contourlets and curvelets is that the contourlet transform is directly defined on digital friendly discrete rectangular grids. Unfortunately, contourlet functions have less clear directional features than curvelets leading to artifacts in denoising and compression. Recently, Lu and Do (2007) have introduced *surfacelets* as the 3D extensions of the 2D contourlets which are obtained by a higher dimensional directional filter bank and a multiscale pyramid. They can be used efficiently to capture and represent surface-like singularities in multidimensional volumetric data involving biomedical imaging, seismic imaging, video processing and computer vision.

The *shearlets*, introduced by Kutyniok and their collaborators (2009), provide an alternative approach to the curvelets and exhibit some very distinctive features. One of the distinctive features of shearlets is the use of shearing to control directional selectivity, in contrast to rotation used by curvelets. Secondly, unlike the curvelets, the shearlets form an affine system. That is, they are generated by dilating and translating one single generating function where the dilation matrix is the product of a parabolic scaling matrix and a shear matrix. In particular, the shearlets can be regarded as coherent states arising from a unitary representation of a particular locally compact group, called the *shearlet group*. This allows one to employ the theory of uncertainty principles to study the accuracy of the shearlet parameters (see Dahlke et al. 2008). Another consequence of the group structure of the shearlets is that they are associated with a generalized multiresolution analysis, and this is particularly useful in both their theoretical and numerical applications (see Kutyniok and Labate 2009; Guo et al. 2006).

For more information about the history of wavelets, the reader is referred to Debnath (1998c).

1.5 Wavelet Bases and Multiresolution Analysis

In the late 1986, Meyer and Mallat recognized that construction of different wavelet bases can be realized by the so-called multiresolution analysis. This is essentially a framework in which functions $f \in L^2(\mathbb{R}^d)$ be treated as a limit of successive approximations $f = \lim_{m \to \infty} P_m f$, where the different $P_m f$ for $m \in \mathbb{Z}$ correspond to smoothed versions of f with a smoothing-out action radius of the order 2^m. The wavelet coefficients $\langle f, \psi_{m,n} \rangle$ for a fixed m then correspond to the difference between the two successive approximations $P_{m-1} f$ and $P_m f$. In the late 1980s, efforts for construction of orthonormal wavelet bases continued rapidly. Battle (1987) and Lemarié (1988, 1989) independently constructed spline orthonormal wavelet bases with exponential decay properties. At the same time, Tchamitchan (1987) gave a first example of biorthogonal wavelet bases. These different orthonormal wavelet bases have been found to be very useful in applications to signal processing, image processing, computer vision, and quantum field theory.

The construction of a "painless" nonorthogonal wavelet expansion by Daubechies et al. (1986) can be considered one of the major achievements in wavelet analysis. During 1985–1986, further work of Meyer and Lemarié on the first construction of a smooth orthonormal wavelet basis on $\mathbb{R}$ and then on $\mathbb{R}^n$ marked the beginning of their famous contributions to the wavelet theory. Many experts realized the importance of the existence of an orthonormal basis with good time–frequency localization. Particularly, Stéphane Mallat recognized that some QMF play an important role in the construction of orthogonal wavelet bases generalizing the classic Haar system. Lemarié and Meyer (1986) and Mallat (1988, 1989a,b) discovered that orthonormal wavelet bases of compactly supported wavelets could be constructed systematically from a general formalism. Their collaboration culminated with a major discovery by Mallat (1989a,b) of a new formalism, the so-called multiresolution analysis. The concept of multiresolution analysis provided a major role in Mallat's algorithm for the decomposition and reconstruction of an image in his work. The fundamental idea of multiresolution analysis is to represent a function as a limit of successive approximations, each of which is a "smoother" version of the original function. The successive approximations correspond to different resolutions, which leads to the name multiresolution analysis as a formal approach to constructing orthogonal wavelet bases using a definite set of rules and procedures. It also provides the existence of so-called scaling functions and scaling filters which are then used for construction of wavelets and fast numerical algorithms. In applications, it is an effective mathematical framework for hierarchical decomposition of a signal or an image into components of different scales represented by a sequence of function spaces on $\mathbb{R}$. Indeed, Mallat developed a very effective numerical algorithm for multiresolution analysis using wavelets. It was also Mallat who constructed the wavelet decomposition and reconstruction algorithms using the multiresolution analysis. This brilliant work of Mallat has been the major source of many recent new developments in wavelet theory. According to Daubechies (1992), "… The history of the formulation of multiresolution analysis is a beautiful example of applications stimulating theoretical development." While reviewing two books on wavelets in 1993, Meyer made the following statement on wavelets: "Wavelets are without doubt an exciting and intuitive concept. The concept brings with it a new way of thinking, which is absolutely essential and was entirely missing in previously existing algorithms."

Each multiresolution analysis determines a scaling function ϕ, that is, a resolution of the so-called dilation equation in the form

$$\phi(x) = \sum_{n=-\infty}^{\infty} c_n \phi(2x - n), \tag{1.5.1}$$

where $\phi \in L^2(\mathbb{R})$. The coefficients $\{c_n\}$ are square-summable real or complex numbers. This scaling function then determines a wavelet

$$\psi(x) = \sum_{n=-\infty}^{\infty} (-1)^n c_n \, \phi(2x - n), \tag{1.5.2}$$

such that the collection $\{\psi_{m,n}\}_{m,n\in\mathbb{Z}}$ with $\psi_{m,n}(x) = 2^{m/2}\psi\big(2^m x - n\big)$ forms an orthonormal basis for $L^2(\mathbb{R})$ after suitable normalization of ψ. This *wavelet orthonormal basis* is so formed by dilating and translating a single $L^2(\mathbb{R})$ function, and therefore, the properties of the basis elements are completely determined by the corresponding properties of wavelet. Thus, the dilation equations play a central role in the construction and resulting properties of multiresolution analysis and wavelet orthonormal basis for $L^2(\mathbb{R})$. This connection is the major reason for the recent rapid development of the study of such equations and their solutions. There are also significant applications of dilation equations to other areas, most notably interpolating subdivision schemes.

Wavelet orthonormal bases are important for many reasons including the major reason that it is possible to find ψ which have good localization in both time and frequency. It is found that all wavelets arising from finite-coefficient dilation equations are smooth and compactly supported. Such wavelets are necessarily well localized in the time domain and have good decay in their Fourier transforms. Also, it is possible to characterize the exact degree of smoothness of these wavelets which means that we can determine the total number of continuous derivatives and possible Hölder exponent of continuity of the last derivative. Daubechies and Lagarias (1992) proved that compact support for wavelets is incompatible with infinite differentiability.

In order to ensure compact support for the wavelet ψ, we assume that the number of nonzero coefficients c_n in (1.5.1) is finite. By translating the scaling function ϕ if necessary, we assume that the dilation equation has only finite number of terms, that is, it has the form

$$\phi(x) = \sum_{n=0}^{N} c_n \phi(2x - n). \tag{1.5.3}$$

That is, we assume $c_n = 0$, for $n < 0$ or $n > N$. We seek square-integrable, compactly supported solutions of Eq. (1.5.3). Such scaling functions are necessarily integrable. It can be shown that compactly supported, integrable solutions of (1.5.3) are unique up to a multiplicative constant. If ϕ is such a compactly supported scaling function, then the associated wavelet ψ is obtained from ϕ by a finite series. Thus, the smoothness of ψ is completely determined by the corresponding smoothness of the scaling function ϕ. Daubechies (1988b) first constructed compactly supported, smooth wavelet which is one of her remarkable contributions to the theory of wavelets.

Without any restrictions on the coefficients $c_0, c_1, \ldots, c_N$, it is possible to determine smooth, compactly supported scaling functions from the dilation equations without regard to their applicability to multiresolution analysis. There are several different methods for constructing such solutions which include Cascade Algorithm method, Fourier transform method, and dyadic interpolation method. These methods are also applicable to dilation equations with integer scale factors other than two, and to some higher dimensional dilation equations.

Inspired by the work of Meyer and stimulated by the exciting developments in wavelets, Daubechies (1988a,b, 1990) made a new remarkable contribution to wavelet theory and its applications. The combined influence of Mallat's work and Burt and Adelson (1983a, l983b) pyramid algorithm used in image analysis led to her major construction of an orthonormal wavelet basis of compact support. Her 1988b paper, dealing with the construction of the first orthonormal basis of continuous, compactly supported wavelets for $L^2(\mathbb{R})$ with some degree of smoothness, produced a tremendous positive impact on the study of wavelets and their diverse applications. Her discovery of an orthonormal basis for $L^2(\mathbb{R})$ of the form $2^{m/2}\psi_r(2^m t - n), m, n \in \mathbb{Z}$, with the support of ψ_r in the interval $[0, 2r + 1]$, created a lot of excitement in the study of wavelets. If $r = 0$, Daubechies' result reduces to the Haar system. This work explained the significant connection between the continuous wavelet on $\mathbb{R}$ and the discrete wavelets on $\mathbb{Z}$ and $\mathbb{Z}_N$, where the latter have become extremely useful for digital signal analysis. Although the concept of frame was introduced by others, Daubechies et al. (1986) successfully computed numerical estimates for the frame bounds for a wide variety of wavelets. In spite of the tremendous success, it is not easy to construct wavelets that are symmetric, orthogonal and compactly supported. In order to handle this problem, Cohen et al. (1992) investigated biorthogonal wavelets in some detail. They have shown that these wavelets have analytic representations with compact support. The dual wavelets do not have analytic representations, but they do have compact support.

In 1990s, another class of wavelets, *semiorthogonal wavelets*, have received some attention. These represent a class of wavelets which are orthogonal at different scales and, for wavelets with nonoverlapping support, at the same scale. Chui and Wang (1991, 1992) and Micchelli (1991) independently studied semiorthogonal wavelets. The former authors constructed B-spline wavelets using linear splines. Then, they used the B-spline wavelets without orthogonalization to construct the semiorthogonal B-spline wavelets. On the other hand, Battle (1987) orthogonalized the B-spline and used these scaling functions to construct orthogonal wavelets. Thus, the difference between Chui and Wang's and Battle's constructions lies in the orthogonal property of the scaling function.

One of the major difficulties with compactly supported orthonormal wavelets is that they lack symmetry. This means that the processing filters are non-symmetric and do not possess a linear phase property. Lacking this property results in severe undesirable phase distortion in signal processing. On the other hand, the semi-orthogonal wavelets are symmetric but suffer from the draw-back that their duals do not have compact support. This is also undesirable since truncation of the filter coefficients is necessary for real-time processing. Cohen et al. (1992) introduced the *biorthogonal multiresolution analysis* in order to produce linear phased finite impulse response filters adapted to the fast wavelet transform. Since for the wavelet transform, linear phase corresponds to a symmetric scaling function, while finite impulse response corresponds to a compactly supported scaling function; Daubechies (1988b) proved that the Haar function $\phi = \chi_{[0,1]}$ is the only orthonormal compactly supported scaling function to be symmetric, so that the orthonormality has to be dropped if linear phase is to be used. Similar results were obtained independently by Vetterli and Herley (1992), they presented a treatment from the "filter design" point of view.

It is well known that the classical orthonormal wavelet bases have poor frequency localization. For example, if the wavelet ψ is band limited, then the measure of the supp of $(\psi_{j,k})^\wedge$ is 2^j-times that of supp $\hat{\psi}$. To overcome this disadvantage, Coifman et al. (1990) constructed univariate orthogonal wavelet packets. Wavelet packets are particular linear combinations or superpositions of wavelets. They are organized naturally into collections, and each collection is an orthogonal basis for $L^2(\mathbb{R})$. Well-known Daubechies orthogonal wavelets are a special case of wavelet packets. Wavelet packets form bases which retain many of the orthogonality, smoothness and localization properties of their parent wavelets, but offer more flexibility than wavelets in representing different types of signals. Wavelet packets, owing to their good properties, have been widely applied to signal processing, coding theory, image compression, fractal theory and solving integral equations, and so on.

The standard construction is to start from a multiresolution analysis and generate the library using the associated quadrature mirror filters. Let $\phi(x)$ and $\psi(x)$ be the scaling function and the wavelet function associated with a multiresolution analysis $\{V_j\}_{j\in\mathbb{Z}}$. Let W_j be the corresponding wavelet subspaces: $W_j = \overline{\text{span}}\{\psi_{j,k} : k \in \mathbb{Z}\}$, where $\psi_{j,k}$ are defined as in (1.4.12). Using the low-pass and high-pass filters associated with the multiresolution analysis, the space W_j can be split into two orthogonal subspaces, each of them can further be split into two parts. Repeating this process j times, W_j is decomposed into 2^j subspaces each generated by integer translates of a single function. If we apply this to each W_j, then the resulting basis of $L^2(\mathbb{R})$ which will consist of integer translates of a countable number of functions, will give a better frequency localization. This basis is called the *wavelet packet basis*. To describe this more formally, we introduce a parameter n to denote the frequency. Set $\omega_0 = \varphi$ and define recursively

$$\omega_{2n}(x) = \sum_{k\in\mathbb{Z}} h_k\, \omega_n(2x-k), \quad \omega_{2n+1}(x) = \sum_{k\in\mathbb{Z}} g_k\, \omega_n(2x-k), \tag{1.5.4}$$

where $\{h_k\}_{k\in\mathbb{Z}}$ and $\{g_k\}_{k\in\mathbb{Z}}$ are the low-pass filter and high-pass filter corresponding to $\phi(x)$ and $\psi(x)$, respectively.

Chui and Li (1993a) generalized the concept of orthogonal wavelet packets to the case of non-orthogonal wavelet packets so that they can be applied to the spline wavelets and so on. The introduction of biorthogonal wavelet packets attributes to Cohen and Daubechies (1993). They have also shown that all the wavelet packets, constructed in this way, are not led to Riesz bases for $L^2(\mathbb{R})$. Shen (1995) generalized the notion of univariate orthogonal wavelet packets to the case of multivariate wavelet packets for the dilation factor $p = 2$, however this construction does not work for $p > 2$. Long and Chen (1997) have reported the non-separable version of wavelet packets on $\mathbb{R}^d$ and generalized the unstability result of non-orthogonal wavelet packets of Cohen–Daubechies to higher dimensional cases. On the other hand, Quak and Weyrich (1997a,b) investigated special type of periodic wavelet packets based on trigonometric polynomial interpolants and studied their decomposition and reconstruction algorithms on closed intervals. The construction of wavelet and wavelet packets related to a class of dilation matrices by the method of unitary extension of a matrix was given by Lian (2004). In his recent paper, Shah (2009) has constructed p-wavelet packets on the positive half-line $\mathbb{R}^+$ using the classical splitting trick of wavelets whereas Shah and Debnath (2011b) have constructed the corresponding p-wavelet frame packets on $\mathbb{R}^+$ using the Walsh–Fourier transform.

Multiwavelets are a natural extension and generalization of traditional wavelets. They have received considerable attention from the wavelet research communities both in the theory as well as in applications. They can be seen as vector valued wavelets that satisfy conditions in which matrices are involved, rather than scalars, as in the wavelet case. Multiwavelets can own symmetry, orthogonality, short support, and high order vanishing moments. However, traditional wavelets cannot possess all these properties at the same time. Multiwavelet system provides perfect reconstruction while preserving length, good performance at boundaries (via linear-phase symmetry), and high order of approximation. In addition, there are more informations of low and high frequency with multiwavelet decomposition than with the traditional wavelets. Multiwavelets have several advantages in comparison with scalar wavelets in image processing and denoising, such as short support, orthogonality, symmetry, and vanishing moments, which are known to be important in image processing and denoising.

The first construction of polynomial multiwavlets was given by Alpert (1993) who used them as a basis for representation of certain operators. Later, Geronimo et al. (1994) constructed two functions $f(t)$ and $g(t)$ whose translations and dilations form an orthonormal basis for $L^2(\mathbb{R})$. The importance for these two functions is that they are continuous, well time-localized (or short support), and of certain symmetry. By imposing the Hermite interpolating conditions, Chui and Li (1996) constructed symmetric antisymmetric orthonormal multiwavelets with particular emphasis on the maximum order of polynomial reproduction and gave examples for length-3 and length-4 multiwavelets. Tham et al. (1998) introduced another class of symmetric-antisymmetric orthonormal multiwavelets which possess a new property called good multifilter properties and demonstrated that they can be useful for image compression. On the other hand, Kessler (2000) has given the general construction of compactly supported orthogonal multiwavelets associated with a class of continuous, orthogonal, compactly supported scaling functions that contain piecewise linear on a uniform triangulation of $\mathbb{R}^2$.

Xia and Suter (1996) introduced *vector-valued multiresolution analysis* and orthogonal vector-valued wavelets. They showed that vector-valued wavelets are a class of generalized multiwavelets and multiwavelets can be generated from the component functions in vector-valued wavelets. Slavakis and Yamada (2001), generalized this concept to biorthogonal matrix-valued wavelets setting. Vector-valued wavelets and multiwavelets are different in the following sense. Vector-valued wavelets can be used to decorrelate a vector-valued signal not only in the time domain but also between components for a fixed time where as multiwavelets focuses only on the decorrelation of signals in time domain. Moreover, prefiltering is usually required for discrete multiwavelet transform but not necessary for discrete vector-valued wavelet transforms. Bacchelli et al. (2002) studied the existence of orthogonal multiple vector-valued wavelets using the subdivision operators where as Fowler and Li (2002) implemented the biorthogonal multiple vector-valued wavelet transforms by virtue of biorthogonal multiwavelets and employed them to study fluid flows in oceanography and aerodynamics. Chen and Cheng (2007) presented the construction of a class of compactly supported orthogonal vector-valued wavelets and investigated the properties of vector-valued wavelet packets. The concept of vector-valued wavelet packets was subsequently generalized to vector-valued multivariate wavelet packets by Chen et al. (2009). In the same year, Xiao-Feng et al. (2009) gave the construction and characterization of all vector-valued multivariate wavelet packets associated with dilation matrix by means of time–frequency analysis, matrix theory, and operator theory.

Wavelet theory has been studied extensively in both theory and applications during the last two decades. This theory has become a promising tool in signal processing, fractals and image processing, and so on, because of their ability to offer good properties like symmetry, certain regularity, continuity, and short support. It is well known that the standard orthogonal wavelets are not suitable for the analysis of high-frequency signals with relatively narrow bandwidth. To overcome this shortcoming, M-band orthonormal wavelets were created as a direct generalization of the 2-band wavelets. The motivation for a larger $M(M > 2)$ comes from the fact that, unlike the standard wavelet decomposition which results in a logarithmic frequency resolution, the M-band decomposition generates a mixture of logarithmic and linear frequency resolution and hence generates a more flexible tiling of the time–frequency plane than that resulting from 2-band wavelet. The other significant difference between 2-band wavelets and M-band wavelets in construction lies in the aspect that the wavelet vectors are not uniquely determined by the scaling vector and the orthonormal bases do not consist of dilated and shifted functions through a single wavelet, but consist of ones by using $M - 1$ wavelets. It is this point that brings more freedoms for optimal wavelet bases.

A function $\phi \in L^2(\mathbb{R})$ is called an *M-band scaling function* if a sequence of closed spaces

$$V_j = \overline{\text{span}}\{M^{j/2}\phi\left(M^j x - k\right), k \in \mathbb{Z}\}, \quad M \geq 2, j \in \mathbb{Z}, \tag{1.5.5}$$

holds the property $V_j \subset V_{j+1}$, $j \in \mathbb{Z}$, and constitutes a multiresolution analysis for $L^2(\mathbb{R})$.

Let $W_j, j \in \mathbb{Z}$ be the direct complementary subspaces of V_j in V_{j+1}. Assume that there exist a set of $M - 1$ functions $\{\psi_1, \psi_2, \ldots, \psi_{M-1}\}$ in $L^2(\mathbb{R})$ such that their translates and dilations form a Riesz basis of W_j, i.e.,

$$W_j = \overline{\text{span}}\left\{M^{j/2}\psi_\ell\left(M^j x - k\right), \ell = 1, 2, \ldots, M - 1, k \in \mathbb{Z}\right\}, \quad j \in \mathbb{Z}. \tag{1.5.6}$$

Then, the functions $\{\psi_1, \psi_2, \ldots, \psi_{M-1}\}$ are called *M-band wavelets*. For more about M-band wavelets and their applications to signal and image processing, we refer to the monograph Sun et al. (2001).

Multiresolution analysis is considered as the heart of wavelet theory. In recent years, there has been a considerable interest in the problem of constructing wavelet bases on various spaces other than $\mathbb{R}$, such as abstract Hilbert spaces, locally compact Abelian groups, Cantor dyadic groups, Vilenkin groups, local fields of positive characteristic, p-adic fields, Hyrer-groups, Lie groups, and zero-dimensional groups. In the p-adic setting, the situation is as follows. In 2002, Kozyrev found a compactly supported p-adic wavelet basis for $L^2(\mathbb{Q}_p)$ which is an analog of the Haar basis. These wavelets are of the form

$$\psi^\gamma_{j,a} = p^{\gamma/2}\chi\left(p^{-1}j\,(p^{-\gamma}x - a)\right)\Omega\left(|p^{-\gamma}x - a|_p\right), \quad x \in \mathbb{Q}_p \tag{1.5.7}$$

where $j = 1, \ldots, p-1, \gamma \in \mathbb{Z}, a \in \mathbb{Q}_p/\mathbb{Z}_p$ which is an analogy of the real Haar basis, where $\Omega(t)$ is the characteristic function of the segment $[0, 1] \subset \mathbb{R}$, $\mathbb{Z}_p$ is the ring of p-adic integers, the function $\chi(x)$ is an additive character of the field of p-adic numbers. The above system is generated by dilations and translations of the wavelet functions $\chi\left(p^{-1}jx\right)\Omega\left(|x|_p\right), x \in \mathbb{Q}_p, j = 1, \ldots, p-1$. It appears that these wavelets are eigen functions of some p-adic pseudo differential operators. This property of wavelets may be used to solve p-adic pseudo-differential equations.

Recently, R.L. Benedetto and J.J. Benedetto (2004) developed a wavelet theory for local fields and related groups. They did not develop the MRA approach, their method is based on the theory of wavelet sets. Moreover, they had doubts that an MRA-theory could be developed because discrete subgroups do not exist in $\mathbb{Q}_p$. Since local fields are essentially of two types: zero and positive characteristic. Examples of local fields of characteristic zero include the p-adic field $\mathbb{Q}_p$ where as local fields of positive characteristic are the Cantor dyadic group and the Vilenkin p-groups. Even though the structures and metrics of local fields of zero and positive characteristics are similar, but their wavelet and multiresolution analysis theory are quite different. Khrennikov et al. (2009) introduced the notion of *p-adic multiresolution analysis* on p-adic field $\mathbb{Q}_p$ and constructed a number of scaling functions generating an MRA of $L^2(\mathbb{Q}_p)$. Later on, Albeverio et al. (2010) proved that all these scaling functions lead to the same Haar MRA and that there exist no other orthogonal test scaling functions generating an MRA of $L^2(\mathbb{Q}_p)$ except those described by Khrennikov et al. (2009).

The concept of multiresolution analysis on a local field K of positive characteristic was introduced by Jiang et al. (2004). They pointed out a method for constructing orthogonal wavelets on local field K with a constant generating sequence. Subsequently, the tight wavelet frames on local fields were constructed by Li and Jiang (2008). They have established necessary condition and sufficient conditions for tight wavelet frame on local fields of positive characteristics in frequency domain. Behera and Jahan (2012a,b) have constructed wavelet packets and wavelet frame packets on local field K of positive characteristic and show how to construct an orthonormal basis from a Riesz basis. Further, they have given the characterization of scaling functions associated with given multiresolution analysis of positive characteristic on local field K. Recently, Shah and Debnath (2013) have constructed tight wavelet frames on local field K of positive characteristic by following the procedure of Daubechies et al. (2003) via extension principles. They also provide a sufficient condition for finite number of functions to form a tight wavelet frame and established general principle for constructing tight wavelet frames on local fields.

1.6 Applications of Wavelet Transforms

Both Weierstrass and Riemann constructed famous examples of everywhere continuous and nowhere differentiable functions. So the history of such functions is very old. More recently, Holschneider (1988) and Holschneider and Tchamitchian (1991) have successfully used wavelet analysis to prove non-differentiability of both Weierstrass' and Riemann's functions.

On the other hand, Beylkin et al. (1991) and Beylkin (1992) have successfully applied multiresolution analysis generated by a completely orthogonal scaling function to study a wide variety of integral operators on $L^2(\mathbb{R})$ by a matrix in a wavelet basis. This work culminated with the remarkable discovery of new algorithms in numerical analysis. Consequently, some significant progress has been made in boundary element methods, finite element methods, and numerical solutions of partial differential equations using wavelets. As a natural extension of the wavelet analysis, Coifman et al. (1989, 1992a,b) in collaboration with Meyer and Wickerhauser discovered wavelet packets to design efficient schemes for the representation and compression of acoustic signals and images. Coifman et al. (1989, 1992a,b) also introduced the local sine and cosine transforms and studied their properties. This led them to the construction of a library of orthogonal bases by extending the method of multiresolution decomposition and using the QMF. Coifman et al. (1989) gave elementary proofs of the L^2 boundedness of the Cauchy integral on Lipschitz curves. Recently, there have also been significant applications of wavelet analysis to a variety of problems in diverse fields including mathematics, physics, medicine, computer science, and engineering.

In recent years, there have been many developments and new applications of wavelet analysis for describing complex algebraic functions and analyzing empirical continuous data obtained from many kinds of signals at different scales of resolution. The most widespread application of the wavelet transform so far has been for data compression. This is associated with the fact that the discrete Fourier transform is closely related to subband decomposition. We close this historical introduction by citing some of these applications which include addressing problems in signal processing, computer vision, seismology, turbulence, computer graphics, image processing, structure of galaxies in the universe, digital communication, pattern recognition, approximation theory, quantum optics, biomedical engineering, sampling theory, matrix theory, operator theory, differential equations, numerical analysis, statistics and multiscale segmentation of well logs, natural scenes, and mammalian visual systems. Wavelets allow complex information such as music, speech, images, and patterns to be decomposed into elementary forms, called simple building blocks, at different positions and scales. These building blocks represent a family of wavelets that are generated from a single function called “mother wavelet” by translation and dilation operations. The information is subsequently reconstructed with high precision. In order to describe the present state of wavelet research, Meyer (1993a) wrote as follows:

> Today the boundaries between mathematics and signal and image processing have faded, and mathematics has benefitted from the rediscovery of wavelets by experts from other disciplines. The detour through signal and image processing was the most direct path leading from the Haar basis to Daubechies’s wavelets.