Wavelet Transform

A New Mathematical Microscope

Sachin P Nanavati and Prasanta K Panigrahi

Sachin P Nanavati is currently working on a project at Physical Research Laboratory, Ahmedabad. He enjoys trekking and bird watching.

Prasanta K Panigrahi is currently with the Quantum Information and Quantum Optics Division at Physical Research Laboratory, Ahmedabad and is on leave from School of Physics, University of Hyderabad. His current research interests are in the areas of quantum information, solitons and wavelets.

Keywords
Fourier transform, discrete wavelet transform (DWT), Haar wavelets.

In the last decade, a new *mathematical microscope* has allowed scientists and engineers to view the details of time varying and transient phenomena, in a manner hitherto not possible through conventional *tools*. This invention, which goes by the name of wavelet transform, has created revolutionary changes in the areas of signal processing, image compression, not to speak about the basic sciences. This novel procedure enables one to achieve the so called *time-frequency* localization and *multi-scale resolution*, by suitably focussing and zooming around the neighborhood of one's choice. Wavelets are of very recent origin; their construction, properties and applications are subjects of intense current research. In this article, we explain with illustrations the working of this transform and its advantages vis-a-vis the Fourier transform. In two companion articles, we describe the procedure to construct wavelet basis sets and their applications to data analysis and image compression.

Introduction

Reducing objects into their basic constituents has been the preferred route of investigation in science and engineering. At an elementary level, we are taught to analyze vectors in terms of their components. In three dimensional Euclidean space, a vector $\vec{A}$ is decomposed in terms of its components as $\vec{A} = A_x\hat{i} + A_y\hat{j} + A_z\hat{k}$. Here, $\hat{i}, \hat{j}$ and $\hat{k}$ are the positive unit vectors in the Cartesian coordinate system, pointing along the orthogonal axes x, y and z, respectively. The fact that these vectors form an *orthonormal set*, *i.e.*,

$$\hat{i}^2 = \hat{j}^2 = \hat{k}^2 = 1 \text{ and } \hat{i}.\hat{j} = \hat{j}.\hat{k} = \hat{k}.\hat{i} = 0, \qquad (1)$$

which is *complete*, allows an arbitrary vector to be expanded in terms of these basic unit constituents. It should be mentioned that any two of the above unit vectors, although orthonormal, do not form a complete basis in three dimensions. Using the above orthonormality conditions, one can find the components of vectors through suitable projections: $A_x = \hat{i}.\vec{A}$, $A_y = \hat{j}.\vec{A}$ and $A_z = \hat{k}.\vec{A}$. Apart from aiding in visualization and facilitating calculations involving vectors, these components can be used to compare two vectors for the purpose of identifying similarities and differences. It is worthwhile to note that these Euclidean unit vectors, by no means, provide an unique basis set. Often, it is convenient to decompose vectors in terms of other unit vectors; for example, circular motion can be better described in a spherical polar coordinate system.

Unlike the constant vector $\vec{A}$ encountered above, the decomposition of a function or a signal $f(t)$, which takes different values at different points of time t, needs more care. Although we have used time t as a continuous variable, one can as well replace it by any other continuous variable like a space coordinate x. If the function is periodic, say with period L $(f(t+L) = f(t))$ and has *finite energy* over one period *i.e.*,

$$\int_{t_0}^{t_0+L} |f(t)|^2 dt < \infty, \qquad (2)$$

then one can decompose the function $f(t)$ in terms of sine and cosine waves, such that

$$f(t) = \frac{a_o}{2} + \sum_{r=1}^{\infty} \left[a_r \cos\left(\frac{2\pi r t}{L}\right) + b_r \sin\left(\frac{2\pi r t}{L}\right) \right] \qquad (3)$$

The well-known Fourier series has found tremendous application in diverse areas, particularly after the discovery of the fast Fourier transform (FFT).

This is the well-known *Fourier series* which has found tremendous application in diverse areas, particularly after the discovery of the fast Fourier transform (FFT). Here a_0, a_r and b_r are constant coefficients representing the average of the function and the amplitudes of the cosine and sine waves, respectively. Much like the unit vectors in Euclidean space, the cosine and sine waves also form an orthogonal basis set, as seen from the following equations:

$$\int_{t_o}^{t_o+L} \sin\left(\frac{2\pi pt}{L}\right) \cos\left(\frac{2\pi rt}{L}\right) dt = 0 \quad \text{for all } p \text{ and } r, \tag{4}$$

$$\int_{t_o}^{t_o+L} \cos\left(\frac{2\pi pt}{L}\right) \cos\left(\frac{2\pi rt}{L}\right) dt = \begin{cases} L & \text{for } p = r = 0, \\ \frac{1}{2}L & \text{for } p = r > 0, \\ 0 & \text{for } p \neq r, \end{cases} \tag{5}$$

$$\int_{t_o}^{t_o+L} \sin\left(\frac{2\pi pt}{L}\right) \sin\left(\frac{2\pi rt}{L}\right) dt = \begin{cases} 0 & \text{for } p = r = 0, \\ \frac{1}{2}L & \text{for } p = r > 0, \\ 0 & \text{for } p \neq r, \end{cases} \tag{6}$$

where integers p and $r \geq 0$. Using the above properties, the coefficients can be easily extracted as

$$a_r = \frac{2}{L}\int_{t_o}^{t_o+L} f(t)\cos\left(\frac{2\pi rt}{L}\right) dt \tag{7}$$

$$b_r = \frac{2}{L}\int_{t_o}^{t_o+L} f(t)\sin\left(\frac{2\pi rt}{L}\right) dt \tag{8}$$

Just as a prism splits white light into its constituent colours, the Fourier transform breaks up a time dependent function into its frequency components. Hence, it is called a *mathematical prism*. To obtain the information about a particular frequency component, we integrate the function $f(t)$, modulated with a cosine or sine function, over one period in the time domain, as in the above expressions. Now the time information of the original function is spread over the entire frequency domain. Practically, it is difficult to retrieve them. For seeing this point clearly, let us imagine doing a Fourier transform of a 60 minute musical concert, where a tuning fork of frequency 440 Hz was played for 20 minutes. Although in the frequency domain, a clear cut peak at 440 Hz will be present, it will be difficult to specify when exactly this tuning fork was played in the time span of the concert. This is because the time information is stored in relative phases (*i.e.*, angles between Fourier coefficients b_r 's and $a_r^{'}s$) of the basis functions. This phase has to be calculated with a precision of $\sim \frac{1}{440\times 20\times 60} = \frac{1}{5{,}28{,}000}$. Computationally, it is impossible to calculate with this required precision! Because of the finite word length of the computer, one can write a real number only up to a certain precision, much lower than the above requirement [1].

Just as a prism splits white light into its constituent colours, the Fourier transform breaks up a time dependent function into its frequency components.

If the function is not periodic but decreases fast enough at infinity, then its Fourier transform can also be defined. It transforms a function $f(t)$, that depends on time, into a new function $\hat{f}(\omega)$ depending on frequency ω which takes continuous values :

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt \tag{9}$$

This new function is called the Fourier transform (FT) of the original function. To obtain the time information of the signal back, the inverse Fourier transform of $\hat{f}(\omega)$ is taken as

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\omega)e^{i\omega t}d\omega \tag{10}$$

One faces difficulties in using the Fourier transform, while dealing with signals or functions which have sharp changes (transient phenomena) or which are rapidly varying in time. This can be better appreciated with the example of the box function and its Fourier transform,

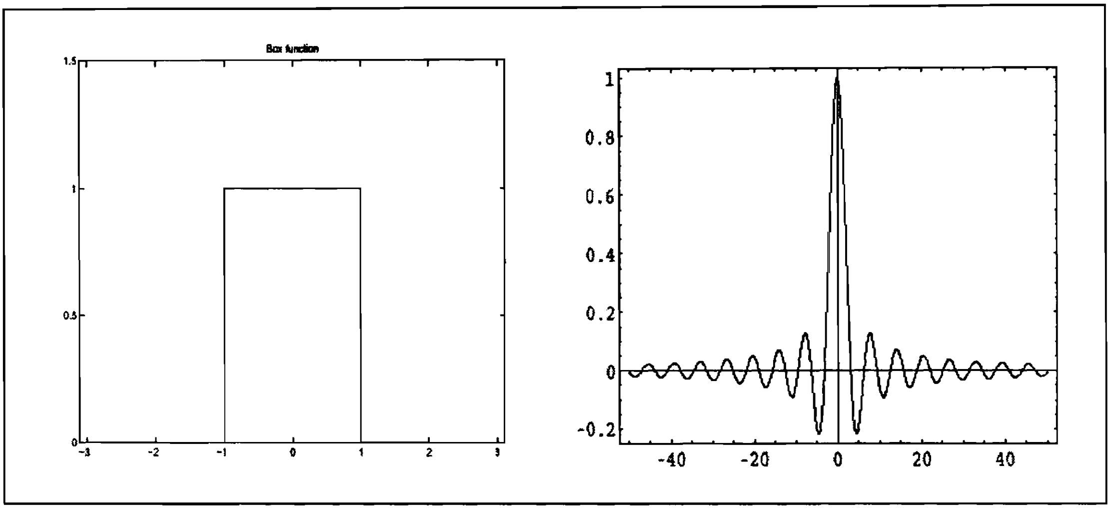

Figure 1. Box function (left) and its Fourier transform, the sinc function (right).

the well-known sinc function ($\frac{\sin\omega}{\omega}$). The box function, as shown in *Figure* 1, has zero value everywhere, except between $t = -1$ and $t = 1$, where it equals unity. One notices immediately that the sinc function is an oscillatory function which decays slowly in the frequency domain, as $|\omega|$ increases. Trying to reconstruct the box function (the inverse transform) after neglecting a few frequency components having very small amplitudes, often necessary in practical applications, leads to distortions at the sharp edges as seen in *Figure* 2. This is

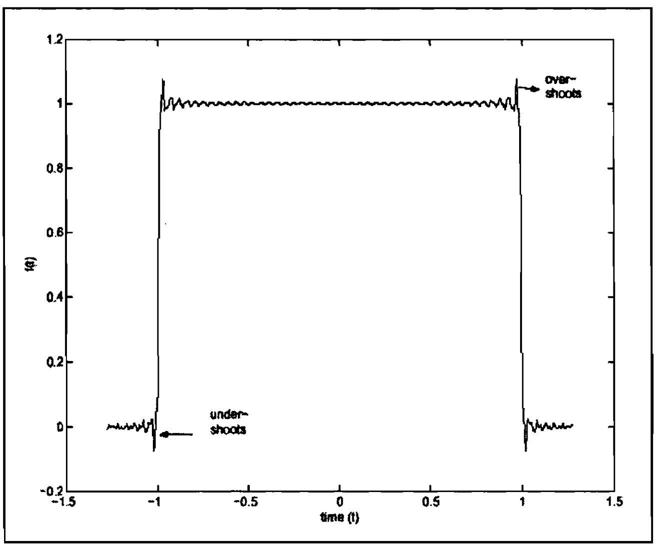

Figure 2. Box function reconstructed after removal of a handful of Fourier coefficients with small values. The overshoots and undershoots at the edges are clearly visible.

the well-known Gibbs' phenomenon [4]. The reason for this difficulty can be explained as follows. To produce the box function, an infinite number of sine and cosine waves, with appropriate amplitudes, are needed to interfere in a manner which is constructive only at the site of the box but destructive elsewhere. It should be noted that the sine and cosine waves themselves extend periodically in the time domain from $-\infty$ to ∞. Hence, any disturbance in this delicate interference, as is done by removing a handful of coefficients having very small values, leads to the overshoots and undershoots at the locations of the sharp changes. One possible solution to this problem is to have a basis set with elements which are themselves localized in time. It will be still better if the basis functions are of a similar form as the function itself; after all, use of one thorn for the removal of another is an ancient wisdom! As we will soon discover, wavelet transform relies precisely on this wisdom.

Use of one thorn for the removal of another is an ancient wisdom! Wavelet transform relies precisely on this wisdom.

Wavelet Transform

A wavelet is a *small wave* which oscillates and decays in the time domain. As discerning readers must have noticed, the logo of *Resonance* resembles a wavelet quite well. Unlike the Fourier transform, wavelets can have infinite varieties which are fundamentally different from each other. The ones which have strictly finite extent in the time domain, are known as *discrete wavelets*, otherwise they go by the name of *continuous wavelets*. Although the most elementary discrete wavelet, named after Haar, was known since 1910, the non-trivial ones and the theory behind wavelet transform are of recent origin. In the following, we will only concentrate on the discrete wavelet transform.

Although the most elementary discrete wavelet, named after Haar, was known since 1910, the non-trivial ones and the theory behind wavelet transform are of recent origin.

A wavelet basis set starts with two orthogonal functions: the *scaling function* or *father wavelet* $\phi(t)$ and the *wavelet function* or *mother wavelet* $\psi(t)$. By scaling and translation of these two orthogonal functions we obtain

A recent use of Haar wavelet was in the extraction of the characteristics of individual thumb impressions by Federal Bureau of Investigation (FBI) of USA.

a complete basis set.

The scaling and wavelet functions, respectively, satisfy

$$\int_{-\infty}^{\infty} \phi(t)dt = A \qquad \text{and} \int_{-\infty}^{\infty} \psi(t)dt = 0 \qquad (11)$$

where A is a constant. The energies of these functions are finite, which means

$$\int_{-\infty}^{\infty} |\phi(t)|^2 dt < \infty \qquad \text{and} \int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty \qquad (12)$$

The scaling function and the mother wavelet are orthogonal to each other:

$$\int_{-\infty}^{\infty} \phi^*(t)\psi(t)dt = 0 \qquad (13)$$

From (11), it is apparent that $\psi(t)$ resembles a wave which is localized in time, in other words it is a small wave or a *wavelet*. In a given wavelet basis set, there is only one scaling function; the rest of the elements are the wavelets. Starting from the mother wavelet, one derives the *thinner* daughter wavelets by appropriate amount of scaling. Scaling is an operation which makes a given object thicker or thinner, by the choice of a parameter. When combined with translation, by amounts commensurate with the size of the wavelets at various scales, one obtains a complete orthogonal basis set, where each element has a finite size. Below we illustrate these points, through the Haar wavelets, the grand old basis.

Haar Wavelet

It is the simplest one to visualize and has found many useful applications. A recent use of Haar wavelet was in the extraction of the characteristics of individual thumb impressions by Federal Bureau of Investigation (FBI) of USA. In this case, the scaling function is given by the box function, while the mother wavelet is an oscillating function of the same maximum height and width. Although we have chosen the height and width as unity here, in principle it is arbitrary, as long as (11-13) are satisfied

It is clear that for the above choice, $\int \phi(t)dt = 1$ and $\int \psi(t)dt = 0$. Also $\int |\phi^2(t)|dt = \int |\psi^2(t)|dt = 1$ and $\int \phi^*(t)\psi(t)dt = 0$. It should be noted that, in the present case, all the basis functions are real. Let us now understand the action of the two operations, translation and scaling

Translation by one unit of the scaling function $\phi(t)$ produces $\phi(t-1)$ which starts at $t = 1$. The reason to choose the translation by unity, same as the width of $\phi(t)$, is to maintain orthogonality between $\phi(t)$ and $\phi(t-1)$. In the same way, one can get $\phi(t-k)$ and $\psi(t-k)$, where k can take integer values, in the range $-\infty$ to ∞. It is easy to convince oneself that all these functions are orthogonal to each other. In the literature, it is customary to denote $\phi(t-k)$ by $\phi_k(t)$ and $\psi(t-k)$ by $\psi_k(t)$.

Although we have now an infinite number of these small building blocks, these are still not complete. For example, $\psi(2t)$ given by an oscillatory function, which takes values $+1$ for $0 \leq t < 1/4$ and -1 for $1/4 \leq t < 1/2$ and its translations, in units of half, are orthogonal to both ϕ_k and ψ_k for any value of k. Note that $\psi(2t)$ is half as thin as $\psi(t)$ and moves in steps half as wide as that of the mother wavelet. The translation step is commensurate with the width. This thinner identical looking wavelet is called a *daughter* wavelet. *Figure* 3 depicts the father, mother, daughter wavelets and their corresponding translations. One can repeat this process to obtain progressively thinner and thinner daughter wavelets, which are orthogonal to the scaling function and all the wavelets prior to them. Note that, in each case, the width of the wavelets gets reduced by a factor of half. All these wavelets can be translated, in their commensurate steps, to cover the entire time axis. The scaling function, the mother wavelet and all these

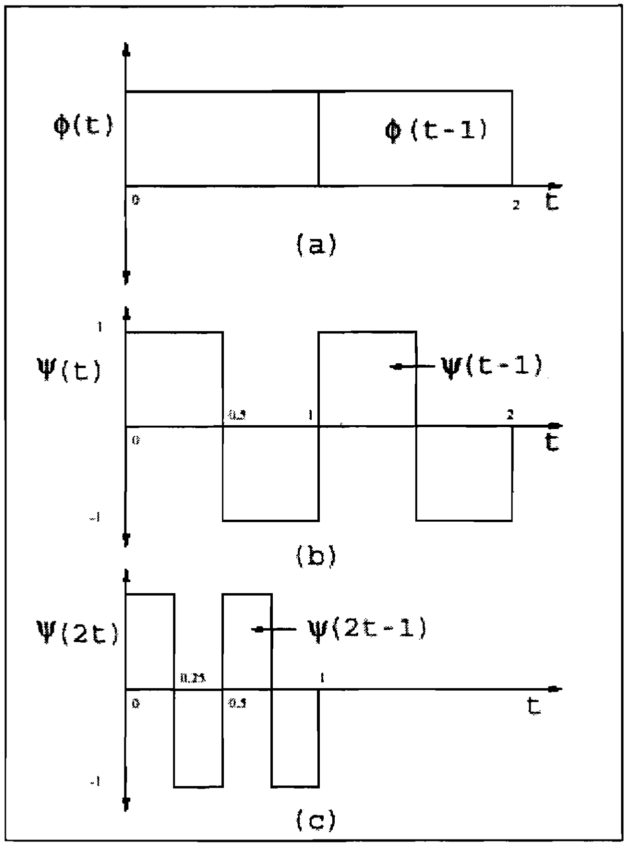

Figure 3. (a) Haar scaling function or the father wavelet $\phi(t)$ $(0 \le t < 1)$ and its translation by one unit $\phi(t-1)$ $(1 \le t < 2)$, (b) mother wavelet function $\psi(t)$ $(0 \le t < 1)$ and its translation by one unit $\psi(t-1)$ $(1 \le t < 2)$ and (c) the scaled mother wavelet, which is a daughter wavelet $\psi(2t)$ $(0 \le t < 0.5)$ and its translation $\psi(2t-1)$ $(0.5 \le t < 1)$.

daughter wavelets (infinite number of them) constitute a complete orthogonal basis set.

As mentioned earlier, the process of making thinner and thicker objects from a given one is known as scaling operation. It should be pointed out that we have taken only one scaling function. It is not difficult to see that $\phi(2t)$, a box function half as wide as $\phi(t)$, is neither orthogonal to $\phi(t)$ nor to $\psi(t)$.

The properly normalized form of the wavelets can be written as

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k) \tag{14}$$

Here, by proper normalization we mean, $\int_{-\infty}^{\infty} |\psi_{j,k}|^2 dt = 1$, for any values of j and k, $2^{j/2}$ being the normalization constant. Note that apart from the translation index k, we have introduced a scaling index j, to properly capture the diadic process of thinning, narrated above in words. From the defining equation of $\psi_{j,k}$, it is clear that $j = 0$ corresponds to the mother wavelet, translated by k units. For values of $j \geq 1$, one obtains the subsequent daughter wavelets. Increasing values of j produce thinner and thinner daughters! We further note that the normalized, thinner daughter wavelet (one corresponding to $j = 1$) is taller by the amount $\sqrt{2}$, as compared to the mother wavelet. The subsequent generations follow the same rule. Hence, as $j \to \infty$, the wavelets become extremely thin with large heights. As will become clear below, the scaling function captures the average of a part of the signal, in an interval determined by its width; the locations of the scaling function window are given by the values of k. The wavelets capture the differences. The mother wavelet obtains the differences of the signal at a scale similar to the scaling function and the thinner daughter wavelets probe the differences at progressively finer scales. This is the reason why wavelet transform is called *multi-resolution analysis*. The fact that the wavelet transform uses a finite size window enables it to capture the local nature of the function or a signal in a much more efficient way than the Fourier transform.

The fact that the wavelet transform uses a finite size window enables it to capture the local nature of the function or a signal in a much more efficient way than the Fourier transform.

Using the complete orthonormal set of basis functions, the wavelet transform of a function $f(t)$ can be written in the form

$$f(t) = \sum_{k=-\infty}^{\infty} c_k \phi_k(t) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} d_{j,k} \psi_{j,k}(t) \qquad (15)$$

Note that, as compared to Fourier transform, discrete wavelet transform involves two indices. Here the coefficients c_k *'s* and $d_{j,k}$ *'s* represent the discrete wavelet transform (DWT) of the function $f(t)$. c_k *'s* capture the average parts, while $d_{j,k}$ *'s* represent the variations, at different scales, present in the function or signal. These features are common to all the wavelets. Explicitly, c_k 's and $d_{j,k}$'s are given, respectively, by $c_k = \int f(t)\phi_k(t)dt$ and $d_{j,k} = \int f(t)\psi_{j,k}(t)dt$.

Let us understand the meaning of these coefficients in the simple Haar basis. Consider c_0, given by $c_0 = \int f(t) \phi(t)dt$; since $\phi(t)$ happens to be the Haar father wavelet, a box function of unit width and height, located at origin, c_0 is the average of the values of the function in the time interval 0 to 1. This represents the average part of the signal in that interval. Similarly, $d_{0,0}$ is the difference in the averages of the function in the two intervals, 0 and 0.5 and 0.5 and 1. Other $d_{j,0}$'s scan the function for still finer variations starting from the origin. The other locations in the time domain are reached by changing values of k.

Because of the presence of two indices j and k, there are several ways to display the wavelet coefficients. We briefly describe here the most popular representation amongst them. Keeping practical applications in mind, instead of a continuous function, we will consider a discrete finite data set $f(t)$, where t takes integral values in a suitable range, for example, $0, 1, 2, 3, \ldots$ First of all, one needs to have $N = 2^n$ data points, n being a positive integer. In case of inadequate data points, there are several methods to augment the data set to the nearest n value, the simplest one being padding with zeros or some other constant numbers. With such a data set, one can go up to maximum of n levels of decomposition. Wavelet coefficients at various levels capture the variations in the function at corresponding scales. As is suggestive, the scaling function coefficients (c_k 's) are called *average or low pass* coefficients and the wavelet coefficients ($d_{j,k}$'s) are called *detail or high pass* coefficients. In any type of wavelet transform, the total number of c_k 's and $d_{j,k}$'s equals the number of data points, as it should be.

Let us now understand the freedom of deciding the number of levels of decomposition. As has been mentioned earlier, the choice of the width of the scaling function is a free parameter, depending upon which the mother and daughter wavelets' sizes are determined. For example, one can take scaling function of width equalling the length of the data set, resulting in only one average coefficient c_0. In case of Haar, this corresponds to the sum of all the data points. The mother wavelet finds the difference between the averages of the first and second halves of the function. The daughter wavelets explore the differences at finer scales, until the maximum resolution is reached due to the discrete nature of the data. In the present case, the thinnest daughter wavelet will find the nearest neighbour differences. In the above, we have performed a full n level decomposition of the data. One could have started with a scaling function having a size half that of the data set, in which case the two average coefficients c_0 and c_1, would have represented the averages of the first and second half respectively. A moment's reflection will reveal that the corresponding two mother wavelet coefficients are equal to the first daughter wavelet coefficients in the earlier case. This example corresponds to a $(n-1)^{th}$ level of decomposition of the data set.

Let us see for ourselves the display of a Haar wavelet transform result of a multibox function, after a one level decomposition. From the above two examples, it is clear that the scaling function now extends only upto two data points. The wavelet and the scaling coefficients are half the size of data points. It is immediately apparent that average coefficients are *similar* to the original data, a feature common to all wavelets. The locations and the degree of variations of the function are present in the detail coefficients, of which only four are non-zero! The differences between the discrete wavelet transform and Fourier transform are clearly *visible* in this example.

Suggested Reading

[1] B B Hubbard, *The World According to Wavelets*, 2nd Edition, Universities Press (India), Hyderabad, 2003.

[2] C S Burrus, R A Gopinath and H Guo, *Introduction to Wavelets and Wavelet Transforms-A Primer*, Prentice-Hall, New Jersey, USA, 1998.

[3] G B Folland, From calculus to wavelets: A New Mathematical Technique, *Resonance*, Vol.2, No.4, pp.25-37, 1997.

[4] S Thangavelu, Fourier Series, *Resonance*, Vol.1, No.10, pp.44-55, 1996.

[5] A Sitaram and S Thangavelu, From Fourier series to Fourier transforms, *Resonance*, Vol.3, No.10, pp.3-5, 1998.

[6] http://www.wavelets.org, Wavelet Digest, a portal which keeps the information of latest happenings in this field. Could be subscribed through mail.

[7] http://engineering. rowan. edu/~polikar/WAVELETS/WTtutorial.html; An introductory tutorial on wavelets by R Polikar.

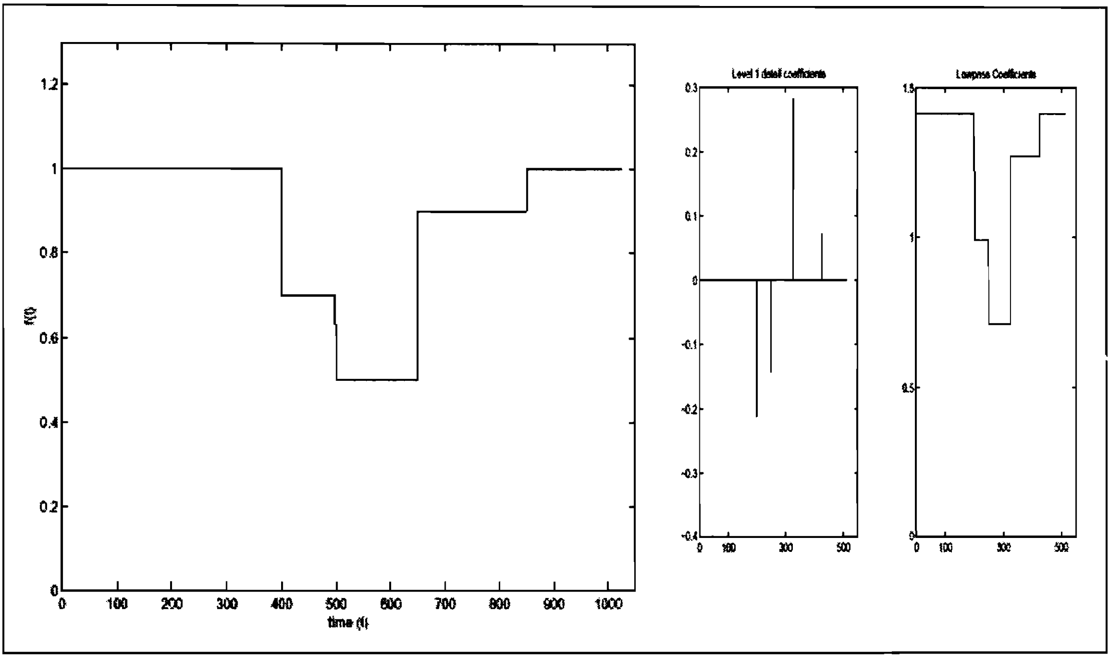

Figure 4. Multi box function (left), with its one level discrete wavelet transform (right).

The enterprising reader can perform a multi-level decomposition and reconstruction, after removal of coefficients having small values, to discover that the problems of overshoots and undershoots plaguing the Fourier transform are absent in discrete wavelet transform.

The enterprising reader can perform a multi-level decomposition and reconstruction to discover that the problems of overshoots and undershoots plaguing the Fourier transform are absent in discrete wavelet transform.

To see the mathematical microscope nature of wavelet transform, let us consider the example of Doppler function (*Figure* 5), containing $N = 2^{11} = 2048$ data points, where variations at different scales and locations are present. Although eleven levels of decomposition are possible, we have performed a five level decomposition, with the Haar wavelets. As is familiar by now, the low pass coefficients at the fifth level, containing $\frac{N}{2^5} = 64$ data points, capture the gross features of the Doppler function. The level one detail coefficients zoom on to the finest of the variations present in the initial part of the data; the mother wavelet coefficients (level five detail coefficients), keep track of the broader variations present in the latter part of the data. The other daughter wavelets localize the variations present at scales in

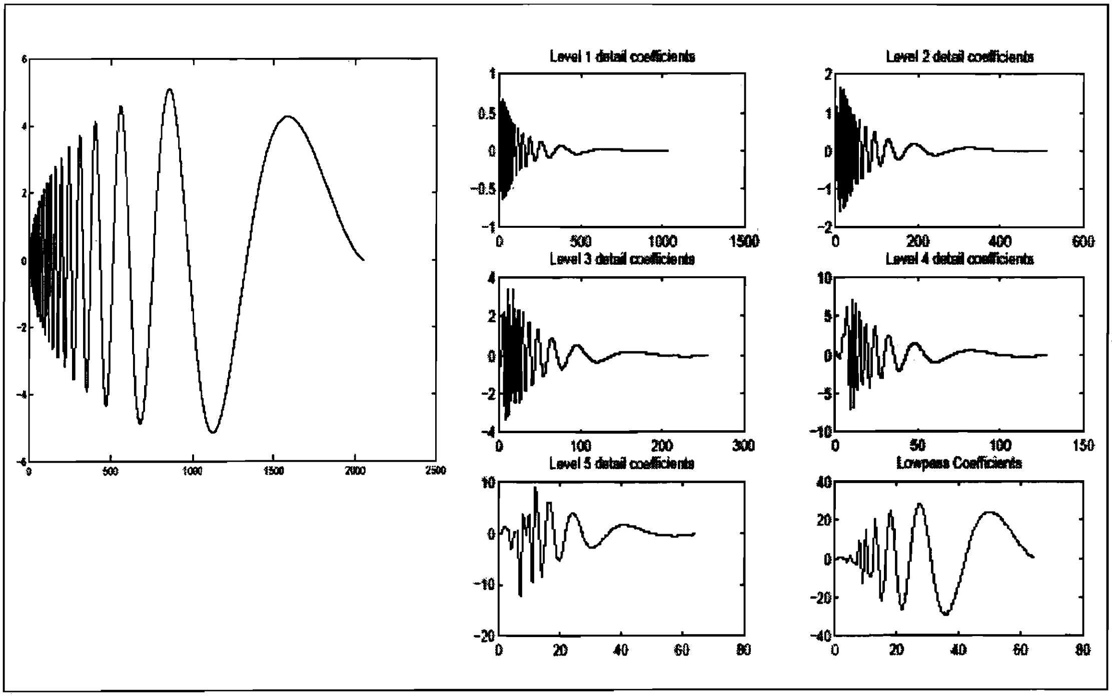

Figure 5. Doppler function (left), with its five level discrete wavelet transform (right).

between the mother and the thinnest daughter wavelet. The neighbourhood to be observed at various levels of resolution is fixed by the size of the scaling function. This is the functioning of the *mathematical microscope.*

The fact that the average coefficients resemble the original function (with lesser data points), makes wavelet transform an ideal tool for a number of applications, a very important one being image compression! Images are two dimensional data sets, analysis of which needs two dimensional wavelet transform. For motivating the readers to a subsequent article, dealing with applications of wavelet transform to image compression, we demonstrate below a one level decomposition of an *academic model*, popular with the workers of wavelet community (*Figure* 6). For decomposition, we have chosen a wavelet from the *Daubechies'* family, the acknowledged queen in the domain of wavelet transform, who in the last decade and a half contributed significantly to the understanding and construction of these *microscopes.*

Figure 6. Original image (left) and its one level, two dimensional discrete wavelet transform (right). The upper left quarter represents the average coefficients while the rest of the three quarters are detail coefficients.

In this article, we have outlined the general structure of wavelet transforms and their usefulness for the analysis of functions or signals with sharp variations. The examples of the functions were chosen so as to bring out the advantages of wavelet transforms over Fourier transform. It should be clearly pointed out that there are other functions for which Fourier transform is eminently more suitable than wavelets. The choice of transforms and the type of wavelet to employ depends on the nature of the data set being studied. We have elaborated on the physically transparent Haar wavelet in this article. In the companion article, the construction and working of other wavelets and their applications will be illustrated. In recent times, wavelet transform has been particularly useful in the study of images and has led to powerful compression techniques. The basic ideas behind the utilization of wavelet transform in image compression will be elaborated in a separate third article.

Address for Correspondence
Sachin P Nanavati and
Prasanta K Panigrahi
Physical Research Laboratory
Navrangpura
Ahmedabad 380 009, India.
Email: spnana@prl.ernet.in
prasanta@prl.ernet.in
URL:www.prl.ernet.in/
~prasanta