

Wavelet Transforms: Application to Data Analysis – I

Jatan K Modi, Sachin P Nanavati, Amit S Phadke and Prasanta K Panigrahi



Jatan K Modi (top left) is currently a visiting faculty at DDIT, Gujarat. His areas of interest are artificial intelligence, compilers and image processing.

Sachin P Nanavati (top right) is currently working on a project at PRL, Ahmedabad. He enjoys trekking and bird watching.

Amit S Phadke (bottom left) is presently working in 'Cytotec', a software firm based in Vadodara, dealing with image processing and 'Flow Cytometry' software. His research interests include image processing and artificial intelligence.

Prasanta K Panigrahi (bottom right) is currently with the Quantum Information and Quantum Optics Division at PRL. His current research interests are in the areas of quantum information, solitons and wavelets.

In this article, we elaborate upon the key ideas underlying the construction of various wavelet basis sets. The roles of translation and scaling, which enable the wavelets to localize variations at desired levels of resolution, are clearly brought out. After explaining the implementation of one dimensional wavelet transform, we illustrate their usefulness through the analysis of a data set.

Introduction

The much-studied Fourier series makes use of the periodic sinusoidal functions, to represent a time dependent function or signal in the frequency domain:

$$f(t) = \frac{a_0}{2} + \sum_{r=1}^{\infty} \left[a_r \cos\left(\frac{2\pi r t}{L}\right) + b_r \sin\left(\frac{2\pi r t}{L}\right) \right] \quad (1)$$

whose continuous version is given by

$$f(t) = \frac{1}{2\pi} \int d\omega F(\omega) \exp(i\omega t). \quad (2)$$

Here a_0 , a_r and b_r are constant coefficients, which represent the average of the function and the amplitudes of the cosine and sine waves with frequency r , respectively. $F(\omega)$ is the Fourier transform of the function $f(t)$, where ω ranges from $-\infty$ to ∞ . At a pictorial level, this means that the function $f(t)$ has been built by the linear superposition of the basic building blocks, the sine and cosine waves, for the Fourier series and plane wave ($\exp(i\omega t)$) for the continuous Fourier transform. The fact that, the plane waves extend from $-\infty$ to ∞ makes them unsuitable to capture local behavior in a function. It should be kept in mind that, in practical applications $f(t)$ is

not continuous, but a discretely sampled data set. As has been explained in the previous article [1], this fact comes out quite clearly in the example of Gibbs' phenomenon. To describe a transient or sharply changing phenomenon by Fourier transform, one needs to keep track of an infinite number of coefficients a_r 's and b_r 's or $F(\omega)$, a practically impossible proposition!

Wavelets in Action

In recent times, wavelet transform has emerged as an advantageous tool for studying time varying and transient phenomena. In a number of cases, they complement the Fourier transform, by providing a better view of the structures present at different scales, as also achieving the so-called *time-frequency* localization, not possible in Fourier domain. Broadly speaking, two different features common to all wavelets are responsible for their utility value. The basis functions of the wavelets are produced from two units, the father wavelet (or scaling function) $\phi(t)$ and the mother wavelet $\psi(t)$, by the operation of *scaling* and *translation*. Depending on the extent of the basis functions, there are two types of wavelets: discrete and continuous. For discrete wavelets, the basis functions have strictly finite extent, which does not hold for the continuous ones. In what follows, we will be primarily dealing with discrete wavelet transform, where the initial size of the father wavelet and its orthogonal mother wavelet is a free parameter, to be chosen depending on the problem at hand.

Daughter wavelets, the other members of the orthogonal basis set, are produced by scaling of the mother wavelets, such that they are orthogonal to the mother and father wavelets and all the other daughter wavelets, preceding them. These are then translated in steps commensurate with their size, so as to provide a complete basis set covering the entire time domain. Hence, the two basic operations scaling and translation, are all that

The basis functions of the wavelets are produced from two units, the father wavelet (or scaling function) $\phi(t)$ and the mother wavelet $\psi(t)$, by the operation of scaling and translation.

Keywords.

Discrete wavelet transform, multi-resolution analysis (MRA), translation, scaling, Daubechies' basis sets.



The mother wavelets capture the variations at a broader scale, whereas the daughter wavelets zoom on to find the differences at progressively finer and finer scales. This is done in a systematic manner, until one reaches the maximum resolution possible in a given data set.

are needed to manufacture a complete orthonormal basis set. The father and mother wavelets need to satisfy the following conditions:

$$\int_{-\infty}^{\infty} \phi(t) dt = A, \quad \int_{-\infty}^{\infty} \psi(t) dt = 0,$$

$$\text{and } \int_{-\infty}^{\infty} \phi^*(t) \psi(t) dt = 0, \quad (3)$$

where A is an arbitrary constant. The energy of these functions is finite, which means

$$\int_{-\infty}^{\infty} |\phi(t)|^2 dt < \infty \quad \text{and} \quad \int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty. \quad (4)$$

In what follows, we normalize both these functions to unity. It should be pointed out that, although we have given a general definition, most of the discrete wavelets are real valued. The scaling function captures the average behavior of the data set under consideration and the wavelets detect the differences (more specifically, the *weighted* averages and differences). These are represented as low-pass (or average) and high-pass (or detail) coefficients of the wavelet transform, respectively. The mother wavelets capture the variations at a broader scale, whereas the daughter wavelets zoom on to find the differences at progressively finer and finer scales. This is done in a systematic manner, until one reaches the maximum resolution possible in a given data set. At each scale, translation allows one to localize the variations in the time domain. In a more precise sense, the basis set is labelled by two parameters j and k , representing scaling and translation indices, respectively. Hence, the normalized basis elements are given by

$$\phi_k(t) = \phi(t - k), \quad \psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k). \quad (5)$$

Here, j and k take integral values, the values of j range from 0 to ∞ , whereas translation index k takes values from $-\infty$ to ∞ . In this notation, the original scaling function $\phi(t)$ corresponds to $\phi_0(t)$ and the mother



wavelet $\psi(t)$ corresponds to $\psi_{0,0}(t)$. As we mentioned earlier, the initial scale is arbitrary. One could have started from the scale index $j = j_0$ instead of $j = 0$, as has been done here. It is worth noting that, the translation step of the daughter wavelet at scale j is $1/2^j$; the thinner daughter wavelets sample the signal at a finer scale, through smaller translation steps. It is not difficult to convince oneself that, in the limit $j \rightarrow \infty$, the wavelet basis forms a complete set [2].

Wavelet Transform

In terms of the scaling and wavelet functions, one can expand a signal or function, which is square-integrable (belonging to L^2 space), in the form

$$f(t) = \sum_{k=-\infty}^{\infty} c(k) \phi_k(t) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} d_j(k) \psi_{j,k}(t). \quad (6)$$

Here, the scaling function and wavelet coefficients are respectively given by,

$$c(k) = \int_{-\infty}^{\infty} \phi_k^*(t) f(t) dt, \quad (7)$$

$$d_j(k) = \int_{-\infty}^{\infty} \psi_{j,k}^*(t) f(t) dt. \quad (8)$$

Note that, for convenience we have denoted the low-pass coefficients as $c(k)$ instead of the mathematically correct form, $c_0(k)$, since by choice the initial scale j has been taken as 0. So far, we have not mentioned about the explicit nature or functional form of the members of the wavelet basis set. As we will soon see, one can find $c(k)$'s and $d_j(k)$'s and also perform the inverse reconstruction operation (*i.e.*, retrieving back $f(t)$ from $c(k)$'s and $d_j(k)$'s), without explicitly knowing the form of the scaling and wavelet functions! As a matter of fact, apart from the simplest discrete wavelet Haar basis, one does not know the explicit form of these functions for any other discrete wavelet basis set. These can be approximated, to any desired degree of accuracy, through

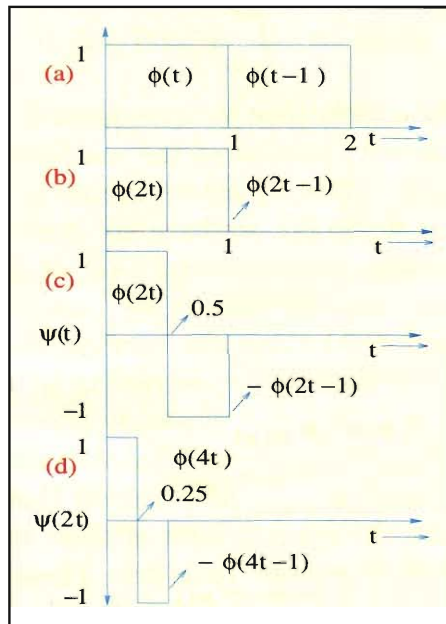
One can find $c(k)$'s and $d_j(k)$'s and also perform the inverse reconstruction operation (*i.e.* retrieving $f(t)$ from $c(k)$'s and $d_j(k)$'s), without explicitly knowing the form of the scaling and wavelet functions!



It should be emphasized that, although the explicit forms of the basis set is not known, there is *no loss* of information in wavelet transforms; the reconstruction of the function $f(t)$ is perfect.

The scaling function plays a key role behind the above miraculous property of the wavelet basis. First of all, the curious reader must have noticed that, there is only one scaling function in any wavelet basis set. The reason for the same is that, unlike the daughter wavelets, a scaling function of higher scale $\phi(2t)$ is neither orthogonal to the previous scaling function nor to the mother wavelet. Interestingly, although $\phi(2t)$ is not a member of a wavelet basis set, it plays a significant role. As an illustration, in case of the Haar basis, it is straightforward to see that $\phi(t) = \phi(2t) + \phi(2t - 1)$ and $\psi(t) = \phi(2t) - \phi(2t - 1)$. *Figure 1* describes this point pictorially; it also shows how the thinner daughter wavelet can be constructed from still thinner scaling functions. In this sense, an appropriate size scaling function can be thought of as a basic building block from which all the wavelets and the father wavelet can be prepared.

Figure 1. Description of multi-resolution analysis (MRA): (a) scaling function $\phi(t)$ and its translation by one unit $\phi(t - 1)$, (b) scaled version $\phi(2t)$ and scaled and translated version $\phi(2t - 1)$, (c) mother wavelet $\psi(t) = \phi(2t) - \phi(2t - 1)$ and (d) a first generation, unnormalized daughter wavelet $\psi(2t) = \phi(4t) - \phi(4t - 1)$.



This fact is known as *multi-resolution analysis (MRA)*, the corresponding mathematical equation goes by the name of *dilation* or *MRA* or *refinement equation*. This is the key equation common to all the wavelets:

$$\begin{aligned} \phi(t) &= \sum_n h(n)\sqrt{2}\phi(2t - n), \\ \psi(t) &= \sum_n \tilde{h}(n)\sqrt{2}\phi(2t - n). \end{aligned} \quad (9)$$

Here, $h(n)$ and $\tilde{h}(n)$ take finite number of values and are respectively known as *low-pass* or *average* and *high-pass* or *detail* filter coefficients. This terminology is derived from the engineering literature, reminding us the fact that many branches of science, engineering and mathematics have contributed to the development of wavelets.

Orthogonality and normalization conditions on scaling and wavelet functions impose a number of conditions on $h(n)$ and $\tilde{h}(n)$:

$$\begin{aligned} \sum_n h(n) &= \sqrt{2}, \quad \sum_n h(n)h(n - 2k) = \delta_{k,0}, \quad \sum_n \tilde{h}(n) = 0, \\ \sum_n \tilde{h}(n)\tilde{h}(n - 2k) &= \delta_{k,0}, \quad \sum_n \tilde{h}(n)h(n - 2k) = 0. \end{aligned} \quad (10)$$

Depending on the length of n , one can explicitly solve these algebraic equations, which characterize and distinguish various basis sets in wavelet transform. For example, when n can take two values, 0 and 1, one finds that,

$$h(0) = \frac{1}{\sqrt{2}}, \quad h(1) = \frac{1}{\sqrt{2}}, \quad \tilde{h}(0) = \frac{1}{\sqrt{2}}, \quad \tilde{h}(1) = \frac{-1}{\sqrt{2}}. \quad (11)$$

Notice that the $1/\sqrt{2}$ factors in the $h(n)$ and $\tilde{h}(n)$ cancel the $\sqrt{2}$ factors in (9) to yield the father and mother wavelets in terms of the scaling function $\phi(2t)$ and $\phi(2t - 1)$. Similarly, if $n = 4$, one finds that,

$$h(0) = \frac{1 - \cos \alpha + \sin \alpha}{2\sqrt{2}} \quad (12)$$

Here, $h(n)$ and $\tilde{h}(n)$ take finite number of values and are respectively known as *low-pass* or *average* and *high-pass* or *detail* filter coefficients. This terminology is derived from the engineering literature, reminding us the fact that many branches of science, engineering and mathematics have contributed to the development of wavelets.



For higher length of n , one gets more freedom in terms of free parameters. Hence, there are an infinite varieties of wavelets and the choice depends on the application at hand.

$$h(1) = \frac{1 + \cos \alpha + \sin \alpha}{2\sqrt{2}} \tag{13}$$

$$h(2) = \frac{1 + \cos \alpha - \sin \alpha}{2\sqrt{2}} \tag{14}$$

$$h(3) = \frac{1 - \cos \alpha - \sin \alpha}{2\sqrt{2}} \tag{15}$$

Here, there is one free parameter α . One gets the Haar coefficients for $\alpha = 0, \frac{\pi}{2}$, and $\frac{3\pi}{2}$. Sometimes, wavelets are made to satisfy additional conditions. For example, a very well-known wavelet, known as *Daubechies'-4* (D4, here 4 indicates the number of filter coefficients) basis set, satisfy $\int_{-\infty}^{\infty} t \psi_{j,k} dt = 0$. With this additional restriction one finds that $\alpha = \frac{\pi}{3}$. The corresponding low-pass filter coefficients, corresponding to $n = 0, 1, 2, 3$ are given by,

$$\begin{aligned} h(0) &= \frac{(1 + \sqrt{3})}{4\sqrt{2}}, & h(1) &= \frac{(3 + \sqrt{3})}{4\sqrt{2}}, \\ h(2) &= \frac{(3 - \sqrt{3})}{4\sqrt{2}}, & h(3) &= \frac{(1 - \sqrt{3})}{4\sqrt{2}}. \end{aligned} \tag{16}$$

The $\tilde{h}(n)$'s follow from the $h(n)$'s, since in general,

$$\tilde{h}(n) = (-1)^n h(n - N + 2k). \tag{17}$$

Here N is an arbitrary integer. For higher length of n , one gets more freedom in terms of free parameters. Hence, there are an infinite varieties of wavelets and the choice depends on the application at hand. The above, apparently abstract condition which the D4 wavelets satisfy, endows it with a magical property. These wavelets, are blind to straight lines! Since, the wavelet coefficients $d_j(k) = \int f(t) \psi_{j,k}(t) dt$, for a straight line, $f(t) = at + b$, these will be zero, as $\int \psi_{j,k}(t) dt = 0$ and $\int t \psi_{j,k}(t) dt = 0$. The information about the straight line, is kept by the low-pass coefficients. In a general signal, variations around a straight line are captured by the



wavelets or high-pass coefficients. By imposing desired *vanishing moments* conditions,

$$\int t^n \psi_{jk}(t) dt = 0, \tag{18}$$

where $n = 1, 2, \dots$ etc, other Daubechies' basis sets can be designed to capture variations around higher polynomial curves. Thus for $n = 2$, we get *Daubechies' 6(D6)* wavelet, whose high-pass coefficients are insensitive to any polynomial of order two or less. If one implements the *vanishing moments* conditions on scaling functions, one gets the wavelet basis sets, called *Coiflets*. One can design various other basis sets, depending upon one's needs.

If one implements the *vanishing moments* conditions on scaling functions, one gets the wavelet basis sets, called *Coiflets*. One can design various other basis sets, depending upon one's needs.

Implementation of Wavelet Transform

The MRA or dilation equation (9), can be used to show that

$$c_j(k) = \sum_n h(n - 2k) c_{j+1}(n), \tag{19}$$

$$d_j(k) = \sum_n \tilde{h}(n - 2k) c_{j+1}(n). \tag{20}$$

As before, n is the length of the filter coefficients. Iteratively, $c_{j+1}(k)$ can be connected to the higher scaling coefficients $c_{j+2}(k)$, which can, in principle, be continued until $j = \infty$. In order to understand the profound implications of the the above results, let us reiterate that,

$$c_{j+1}(k) = \int_{-\infty}^{\infty} \phi_{j+1,k}^*(t) f(t) dt, \tag{21}$$

where the normalized scaling function at scale j is given by

$$\phi_{j,k}(t) = 2^{j/2} \phi(2^j t - k). \tag{22}$$

From above it is clear that, $\phi_{j+1,k}(t)$ is thinner as compared to $\phi_{j,k}(t)$ with an additional normalization factor $\sqrt{2}$. In the limit $j \rightarrow \infty$, the scaling function will become extremely *thin* and *tall*, mimicking a Dirac delta



As has been noted earlier, in practical applications $f(t)$ is not continuous but represents a discretely sampled data set. For this discrete data, the maximum resolution is achieved at a scale value where the scaling function's span is less than the distance between the individual data points.

function; hence, the corresponding low-pass coefficient $c_{j \rightarrow \infty}(k)$ is nothing but the value of the function $f(t)$ at location k . As has been noted earlier, in practical applications $f(t)$ is not continuous but represents a discretely sampled data set. For this discrete data, the maximum resolution is achieved at a scale value where the scaling function's span is less than the distance between the individual data points. This could be achieved by q number of iterations ($\log_2 N$), where the total number of data points, $N = 2^q$. Starting from the highest resolution low-pass coefficients, *i.e.*, the data points, all the other lower scale high-pass and low-pass coefficients could be obtained from equations (19) and (20), by performing desired number of iterations. For example, a single iteration yields level 1 low-pass and high-pass coefficients, each containing half the number of data points of the original signal. These, low-pass coefficients can again be used as a seed to generate, level two low-pass and high-pass coefficients, each containing $\frac{N}{4}$ data points. This procedure can be carried out up to maximum q^{th} level of decomposition, where the low-pass function contains a single data point. Explicitly, for the Haar wavelets, let us illustrate the above procedure by an example. Consider the given signal $f(t)$ as the sequence of points, a_0, a_1, a_2, a_3 . In this case, there are 2^2 data points, hence $q = 2$ is the maximum level of decomposition possible.

As has been seen earlier, for the Haar wavelet, $h(n)$ and $\tilde{h}(n)$ are given by $[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$ and $[\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}]$ respectively. After applying low-pass and high-pass filter coefficients, as given in equations (19), (20), we get $[(\frac{a_0+a_1}{\sqrt{2}}) (\frac{a_2+a_3}{\sqrt{2}})]$ and $[(\frac{a_1-a_0}{\sqrt{2}}) (\frac{a_3-a_2}{\sqrt{2}})]$. These are the level-1 low-pass and high-pass coefficients respectively. Note that we have skipped alternate points while computing low and high-pass coefficients. This process is called down-sampling or decimation by two. The four original data points can be reconstructed from the two, level-1 low-pass and



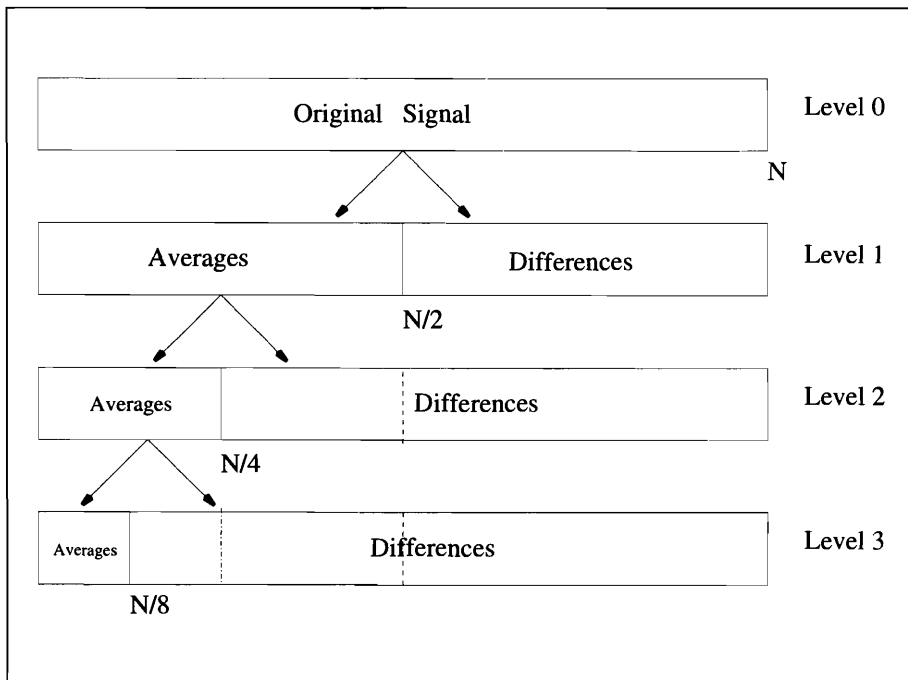
high-pass coefficients, by solving four simultaneous algebraic equations. It is easy to convince oneself that, not skipping the alternate points would have amounted to eight such equations, where four are redundant. Modulo a normalization factor, the first bracket contains nearest neighbour *averages* and the second one consists of nearest neighbour *differences*.

In the second level of decomposition, the differences are kept untouched, while the above procedure is recursively applied on first low-pass coefficients yielding,

$$\frac{1}{\sqrt{2}} \left[\left(\frac{a_0+a_1}{\sqrt{2}} \right) + \left(\frac{a_2+a_3}{\sqrt{2}} \right) \right] \text{ and } \frac{1}{\sqrt{2}} \left[\left(\frac{a_3+a_2}{\sqrt{2}} \right) - \left(\frac{a_1+a_0}{\sqrt{2}} \right) \right].$$

We now have performed the maximum level of decomposition possible, by getting a single low-pass coefficient. The low-pass coefficient is the average of all the data points and the single level-2 high-pass coefficient is the difference of the nearest neighbour averages. A pictorial representation of the above decomposition procedure, upto three levels is given in *Figure 2*.

Figure 2. 1-D, three level, forward discrete wavelet transform (DWT).



A very important property of the wavelet transform is that, it conserves the energy of the function or signal. It is known as Parseval's theorem.

A very important property of the wavelet transform is that, it conserves the energy of the function or signal. It is known as Parseval's theorem. For a discrete signal of length N with the data points a_i 's, the energy is given by

$$E = \sum_{i=0}^{N-1} |a_i|^2. \tag{23}$$

In the wavelet domain, the same is given by

$$E = \sum_{k=0}^{\infty} |c(k)|^2 + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} |d_j(k)|^2. \tag{24}$$

For our above data set, the energy is given by $a_0^2 + a_1^2 + a_2^2 + a_3^2$. After a 2-level decomposition,

$$\begin{aligned} E &= \left[\frac{1}{\sqrt{2}} \left(\frac{a_0 + a_1}{\sqrt{2}} \right) + \frac{1}{\sqrt{2}} \left(\frac{a_2 + a_3}{\sqrt{2}} \right) \right]^2 + \\ &\quad \left[\frac{1}{\sqrt{2}} \left(\frac{a_3 + a_2}{\sqrt{2}} \right) - \frac{1}{\sqrt{2}} \left(\frac{a_1 + a_0}{\sqrt{2}} \right) \right]^2 + \\ &\quad \left[\left(\frac{a_1 - a_0}{\sqrt{2}} \right)^2 + \left(\frac{a_3 - a_2}{\sqrt{2}} \right)^2 \right] \\ &= a_0^2 + a_1^2 + a_2^2 + a_3^2. \end{aligned} \tag{25}$$

The necessity of the $\frac{1}{\sqrt{2}}$ normalization factors at each level, is now clear. Without these factors, the energy will not be conserved! Depending on the wavelet basis, these factors change, however energy remains conserved in all the basis sets. Let us see the 1-dimensional wavelet analysis of a more realistic data set consisting of 1024 (2^{10}) points. In principle, 10 levels of decomposition are possible, one may choose a lower level of decomposition. In *Figure 3* we display the data and in *Figure 4* its wavelet decomposition up to 4 levels.

The level-1 high-pass coefficients are 512 in number, while the level-4 low-pass coefficients number 64. The resemblance of the low-pass coefficients with the original



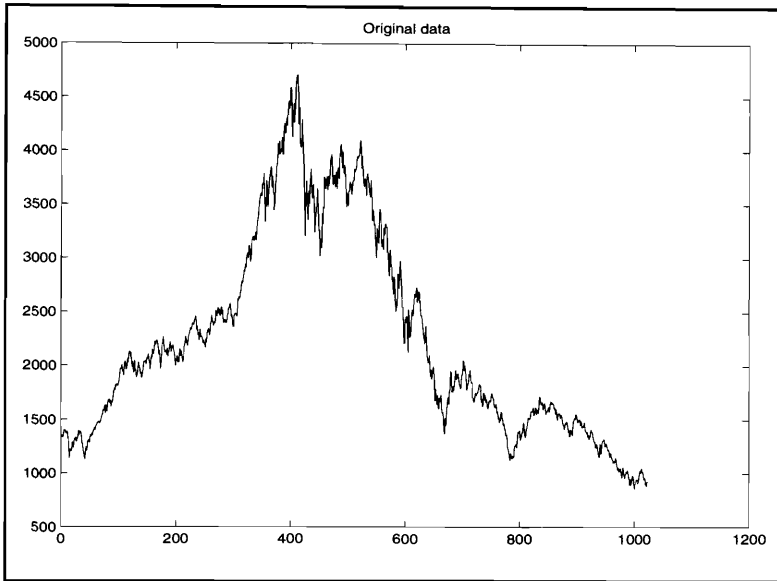


Figure 3. Original data (1024 data points)

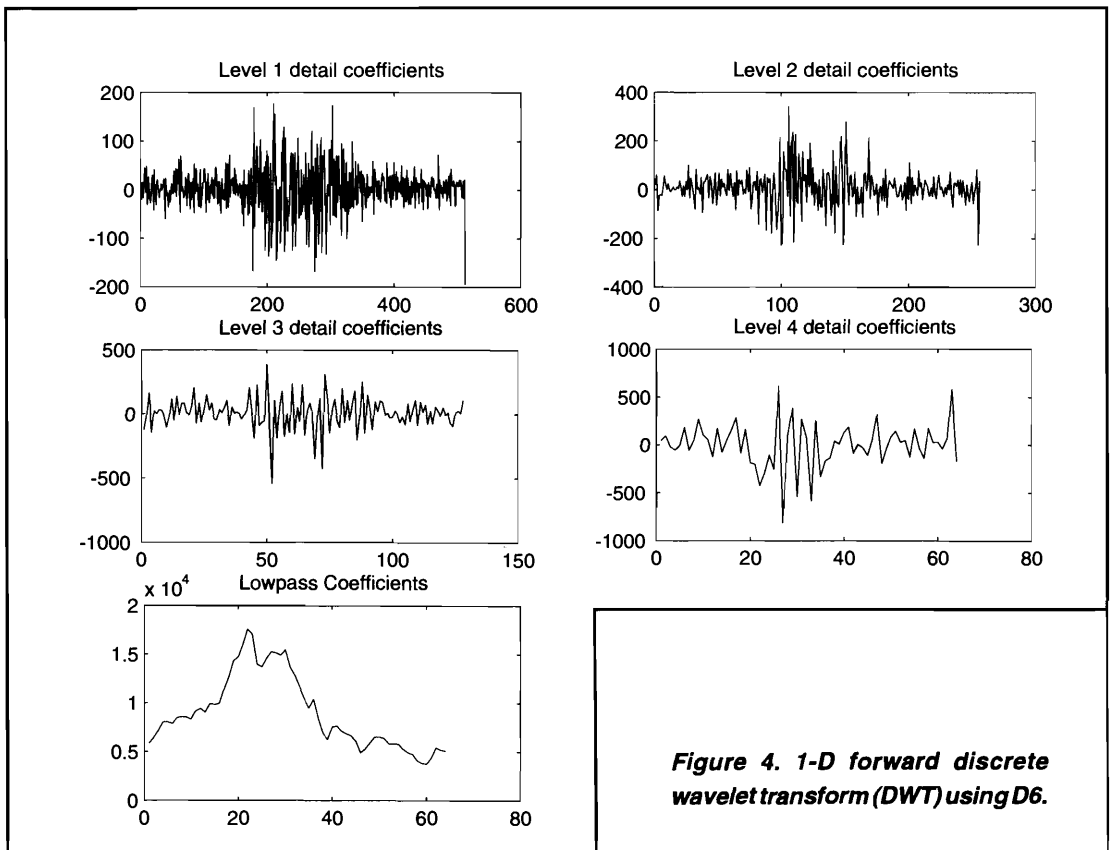


Figure 4. 1-D forward discrete wavelet transform (DWT) using D6.



data is evident; this feature is common to all wavelets. We have chosen D6 wavelet for analysis. One notices fluctuations at various scales in the wavelet domain. Certain prominent variations are clearly visible. One also finds coefficients having large values at the end of the detail coefficients, whereas significant variations are absent in the corresponding location of the data set. This is an artifact of the wavelet transform, arising because of our use of the periodic boundary condition.

It is sufficient to add, without going into details, that wavelets have found applications in the analysis of data sets derived from diverse sources like stock market, cosmic rays, genome project, musical scores, seismic waves, etc. The power to disentangle variations at different scales and also to localize them in an optimal manner, has made these tiny waves create ripples in diverse and otherwise unrelated areas.

Address for Correspondence

Jatan K Modi and
Amit S Phadke
Dharmsinh Desai Institute of
Technology, Nadiad 387 001,
India.

Sachin P Nanavati
Physical Research Laboratory
Navrangpura
Ahmedabad 380 009, India.
Email: sponana@prl.ernet.in

Prasanta K Panigrahi
National PARAM Supercon-
ducting Facility, Centre for
Development of Advanced
Computing (C-DAC), Pune
University Campus, Ganesh
Khind, Pune 411 007, India.
Email: prasanta@prl.ernet.in
URL: www.prl.ernet.in/
~prasanta

Suggested Reading

- [1] S Nanavati and P Panigrahi, *Wavelet transform: A new mathematical microscope*, *Resonance*, Vol.9, No.3, pp 50-64, March 2004.
- [2] I Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, USA. 1992.
- [3] B B Hubbard, *The World According to Wavelets*, 2nd edition, Universities Press (India), Hyderabad, 2003.
- [4] G B Folland, *From Calculus to Wavelets: A New Mathematical Technique*, *Resonance*, Vol.2, No.4, pp.25-37, April 1997.
Wavelet Information on Internet (as on March 2004)
- [5] <http://www.wavelets.org>, *Wavelet Digest*, a portal which keeps the information of latest happenings in this field. Could be subscribed through mail.
- [6] <http://www.public.iastate.edu/~rpolikar/WAVELETS/WTtutorial.html>, *A Introductory Tutorial on Wavelets by Robi Poliker*.