

From Fourier Transform to Wavelets

Otto Seppälä
April 2001

1. TRANSFORMS	2
1.1. BASIS FUNCTIONS	2
1.1.1. SOME POSSIBLE BASIS FUNCTION CONDITIONS	2
1.1.1.1. Orthogonality	2
1.1.1.2. Redundancy	2
1.1.1.3. Compact support	2
1.2. FOURIER TRANSFORMS	3
1.2.1. DISCRETE FOURIER TRANSFORM	3
1.2.2. SHORT TERM FOURIER TRANSFORM	4
1.2.3. CAN IT BE FIXED?	4
2. WAVELETS	5
2.1. BRIEF HISTORY OF WAVELETS	5
2.2. CONTINUOUS WAVELET TRANSFORM	6
2.2.1. WAVELET REQUIREMENTS	6
2.2.1.1. Admissibility	6
2.2.1.2. Regularity	6
2.3. DISCRETE WAVELET TRANSFORM	7
2.3.1. PROBLEMS WITH CWT	7
? Redundancy	7
? Infinite solution space	7
? Efficiency	7
2.3.2. WAVELET DISCRETIZATION	7
2.3.2.1. Wavelet series decomposition	8
2.3.2.2. Wavelets as a band-pass filter bank	8
2.3.3. SUBBAND CODING AND MULTIREOLUTION ANALYSIS	9
3. REFERENCES	10

1. Transforms

When data of any kind is recorded, it hardly ever reveals its true nature from the point you are standing at. Transforms made on the data perform the task of changing this standpoint. A basic transform in our physical world is a transform from the time-domain to the frequency-domain and vice versa. Our ears for example are made to find out the frequencies and phases of the sound signal. The frequencies are used to understand the “message”, the phases carry information which is mostly used for spatial hearing (low frequencies only). It is clear that such transforms have a high analytical value.

1.1. Basis Functions

The idea of a having basis function is that any signal can be described as a weighed sum of a family of functions. These functions are the basis functions of the transform. Transforms are often just changes from one basis function to another, although other transform types exist. For example in cryptology it is not practical to have a “codebook”, but rather use other mathematical methods for performing the transform.

1.1.1. Some possible basis function conditions

Depending on the features of the transform, the basis functions have to fulfill different requirements. These conditions allow for a number of nice mathematical tricks. Some conditions are presented here a couple of other ones pop up when we look into the wavelet requirements.

1.1.1.1. Orthogonality

From an analysis point of view, it is practical that a set of functions is orthogonal. For a set of functions $f_1..f_n$, orthogonality on a given interval $a \leq x \leq b$ requires that the correlation between different functions is zero. The δ denotes the Kronecker delta-function.

$$\int_a^b f_n(x) f_m(x) dx = \delta(m-n) \quad (1)$$

Orthogonal basis functions describe the signal explicitly. In other words, there exists only one possible transform for each signal, when such functions are used. Orthogonal basis functions are also a good choice for compression tasks.

1.1.1.2. Redundancy

Not all transforms aim on analyzing the signal afterwards. If we want to protect the data against errors, it might be rational to add redundancy. In these cases, the choice of basis functions leads to a redundant set, where only certain basis functions exist, and others are “forbidden”. In case of an error, the basis function can be corrected to the nearest “legal” one according to the distance metric of the function space.

1.1.1.3. Compact support

Compact support means that wherever the function is not defined it will have a value of zero. In numerical calculation, this is an advantage. It is logical that all known zeroes speed things up as we can use algorithms that exploit this feature. Compact support is an important requirement in wavelet decompositions, because it also tells about the locality of the wavelet in the time domain.

1.2. Fourier transforms

Fourier transform is by all means the most well known frequency domain analysis tool. There still exist certain problems that cannot be elegantly solved by using Fourier transformations. A logical preface to wavelets is to take a look at the different Fourier methods available.

1.2.1. Discrete Fourier Transform

The discrete Fourier transform uses sines and cosines as basis functions. This limits its practical use to periodic, stationary signals. There are also boundary effects as the transform is originally designed for infinite signals. The biggest problem is that frequency components in a signal cannot be localized in time.

The original DFT was slow, but with the internal redundancy taken away, the Fast Fourier Transform is a nice tool for signal analysis. FFT is implemented by exploiting many times the fact that one DFT transform can be separated into two transforms of half the size. This halves the calculation in the 1D transform. A tenfold division into smaller DFTs brings a thousand-fold speedup.

$$\begin{aligned}
 F(u) &= \sum_{x=0}^{N-1} f(x) e^{-j2\pi ux/N} \\
 &= \sum_{x=0}^{N/2-1} f(2x) e^{-j2\pi u(2x)/N} + \sum_{x=0}^{N/2-1} f(2x+1) e^{-j2\pi u(2x+1)/N} \\
 &= \sum_{x=0}^{N/2-1} f(2x) e^{-j2\pi u(2x)/N} + e^{-j2\pi u/N} \sum_{x=0}^{N/2-1} f(2x+1) e^{-j2\pi u(2x)/N} \\
 &= \frac{1}{2} (F_{\text{even}}(u) + e^{-j2\pi u/N} F_{\text{odd}}(u))
 \end{aligned} \tag{2}$$

The multidimensional Fourier transform is also separable to 1D Fourier transforms. The multidimensional signals can first be transformed in the direction of one axis and then the once transformed signal is transformed one by one along each of the axis. This is a significant advantage as the number of calculations goes from $O[N^2]$ down to $O[DN]$.

$$\begin{aligned}
 F(u,v) &= \frac{1}{NM} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) e^{-j2\pi (ux/M + vy/N)} \\
 &= \frac{1}{NM} \sum_{x=0}^{M-1} e^{-j2\pi (ux/M)} \sum_{y=0}^{N-1} f(x,y) e^{-j2\pi (vy/N)}
 \end{aligned} \tag{3}$$

1.2.2. Short Term Fourier Transform

The Short Term Fourier Transform was designed to overcome the problem with localising the frequency components in time. The idea is to use a window to get a part of the signal. Then stationarity is assumed in that little piece of data.

Heisenberg's uncertainty relation states the relation of the window width to the frequency localization capability. This basically means that the more narrow the window is, the more inaccurate the results will be.

Low frequencies cannot be detected with small windows.

A very big problem however is that there exists no inverse transform for the STFT and therefore its suitable for only gathering data etc.

1.2.3. Can it be fixed?

As it will be later shown in this paper, the wavelets use automatically optimal window sizes for different wavelet widths and they are localized in both time and frequency domain. This fixes the localization problem. As they model only small portions of the signal, the signal doesn't have to be stationary. A huge difference to the STFT is that wavelets can be used for decomposition and exact reconstruction of the signal.

2. Wavelets

2.1. Brief History of Wavelets

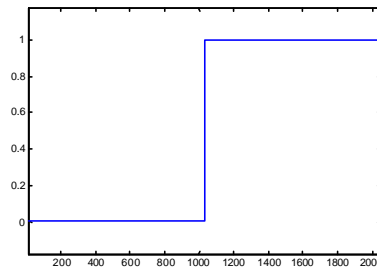


Figure 1 Haar scaling function

- ?? 1909 Alfred Haar introduced the first compactly supported family of functions. These are the Haar basis.
- ?? 1946 Gabor introduces a family of non-orthogonal wavelets with unbounded support. These wavelets are based on translated gaussians.
- ?? 1976 Croisier, Esteban and Galand introduce filter banks for decomposition and reconstruction of a signal. Flanagan, Crochiere and Weber introduce roughly the same idea in speech acoustics, called sub-band coding.
- ?? 1982 Jean Morlet uses Gabor wavelets to model seismic signals.

Morlet Wavelet

$$\psi(t) = \sqrt{2} e^{-t^2/4} (e^{j\pi t} - e^{-j\pi t/4}) \quad (4)$$

- ?? 1987-1993 Stephane Mallat and Yves Meyer created the multiresolution theory which resulted in the Discrete Wavelet Transformation.
- ?? 1988 Ingrid Daubechies developed an orthonormal, compactly supported family of wavelets. These wavelets were created through iterative methods, not as with explicit functions.

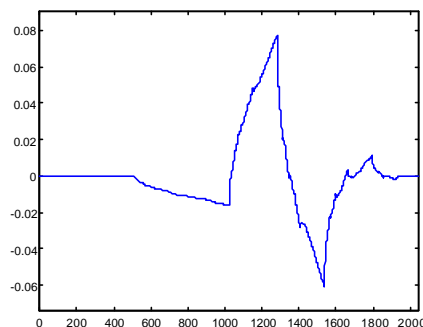


Figure 2. A sample Daubechies Wavelet

- ?? 1996 Sweldens, Calderbank and Daubechies propose the lifting scheme for developing "second-generation" wavelets that can be derived from FIR filters and map integers to integers.

2.2. Continuous Wavelet Transform

The wavelets form a family. The basic form is called the mother wavelet. All the daughter wavelets are derived from this wavelet according to equation (5).

$$\psi_{s,t}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-t_0}{s}\right) \quad (5)$$

The two new variables, s and t are the scale and translation of the daughter wavelet. The term $s^{-1/2}$ normalizes the energy for different scales, whereas the other terms define the width and translation of the wavelet.

The Continuous Wavelet Transform (CWT) is defined as follows. The asterisk denotes a complex conjugate function.

$$W(s, t) = \int_{-\infty}^{\infty} f(\tau) \psi_{s,t}^*(\tau) d\tau \quad (6)$$

The inverse transform in the other hand is :

$$f(t) = \int_0^{\infty} W(s, t) \psi_{s,t}(t) ds \quad (7)$$

A very important fact about wavelets is that the transform itself poses no restrictions whatsoever on the form of the mother wavelet. This is an important difference to other transforms. One can thus design wavelets to suite the requirements of the situation.

2.2.1. Wavelet requirements

There are however two conditions that the wavelet has to fulfil. The admissibility condition (8) and the regularity condition.

2.2.1.1. Admissibility

$$\int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty \quad (8)$$

Two results arise from this condition.

1. The transformation is invertible. (All square integratable functions that satisfy the admissibility condition can be used to analyse and reconstruct any signal.)
2. The function must have a value of zero at zero frequency. Therefore, the wavelet itself has an average value of zero. This means that the wavelet describes an oscillatory signal, where the positive and negative values cancel each other. This gives the 'wave' in wavelet.

2.2.1.2. Regularity

The regularity condition requires that the mother wavelet has to be locally smooth and concentrated in both the time and frequency domains. As a result of this condition we define a new concept, vanishing moments. If the Fourier transform of the wavelet is M times differentiable the wavelet has M vanishing moments.

$$\int_{-\infty}^{\infty} x^m \psi(x) dx = 0 \quad m \in [0, M] \quad (9)$$

2.3. Discrete Wavelet Transform

2.3.1. Problems with CWT

As nice as the theory of the CWT is, it still has three major problems. These problems make the continuous wavelets hard to implement for solving any real problem.

?? Redundancy

The basis functions for the continuous wavelet transform are shifted and scaled versions of each other. It is clear that these cannot form a very orthonormal base.

?? Infinite solution space

The result holds an infinite number of wavelets. This makes it even harder to solve and in the other hand, it makes it hard to find the desired results out of the transformed data.

?? Efficiency

Most of the transforms cannot be solved analytically. Either the solutions have to be calculated numerically, which takes an incredible amount of time, or they can be calculated through optics etc. In order to use the wavelet transform for anything rational, we must find very efficient algorithms.

2.3.2. Wavelet Discretization

The redundancy in the continuous wavelets is can be fixed. First the continuous variables s and t are discretized so that the mother wavelet can only be scaled and translated in discrete steps. The discretized values of t are made dependent on s in such a manner, that low frequency components are sampled less often. As a result of the discretization we get a new equation for deriving daughter wavelets.

$$\psi_{j,k}(t) = \frac{1}{\sqrt{S_0^j}} \psi\left(\frac{t - t_0 S_0^j}{S_0^j}\right) \quad (10)$$

S_0 is usually assigned a value of 2 and t_0 the value of one. This results in dyadic sampling. The picture(1) tries to show, how the samples are located in the discretized time-scale space.

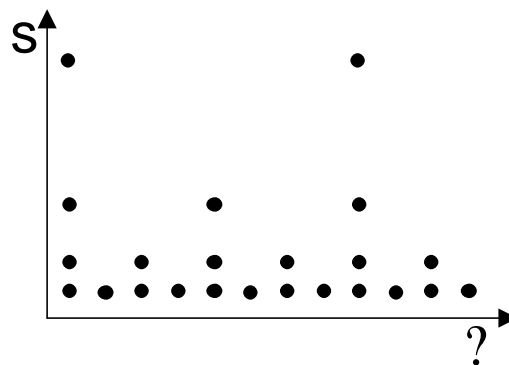


Figure 3. Discretized samples in the time-space domain – Dyadic sampling

2.3.2.1. Wavelet series decomposition

A transform performed with the discrete wavelets is called the wavelet series decomposition. For this transform to be of any major use, the transform must be invertible. It can be proven that if equation (11) is satisfied, the transform is invertible. When $A=B$ the same wavelet can be used to do the transformation in both directions. When the equality holds we have a so called tight frame.

$$A\|f\|^2 = \sum_{j,k} \left| \langle f, \psi_{j,k} \rangle \right|^2 = B\|f\|^2 \quad A = B \quad (11)$$

If we also require that all the discretized wavelets form an orthonormal basis, the following has to be true for all values of j,k,m,n .

$$\int_{-\infty}^{\infty} \psi_{j,k}(t) \psi_{n,m}^*(t) dt = \delta_{(j,n)} \delta_{(k,m)} \quad (12)$$

Meaning that the equation is true only if $j=n$ and $k=m$. Orthogonality is not essential for discrete wavelets to work, and sometimes it is not required, as it reduces the flexibility of the wavelets.

2.3.2.2. Wavelets as a band-pass filter bank

The regularity condition ensures a band-pass like spectrum for the mother wavelet in the frequency domain. The wavelet discretization we made, used scales that were related to powers of s_0 . But what happens when a signal is compressed in time? A familiar result in mathematics and signal processing is shown in equation (13).

$$F\{f(at)\} = \frac{1}{|a|} F\left(\frac{f}{a}\right) \quad (13)$$

This tells that when the signal is compressed to half of its original width, its spectrum is made twice as wide and the components of the spectrum are all transposed to frequencies twice as high as the original. This allows for covering the whole spectrum, because we can design the mother wavelet in such a way that these band-pass filters form a "chain" in which each filter is overlapping a little with the previous and the following one.

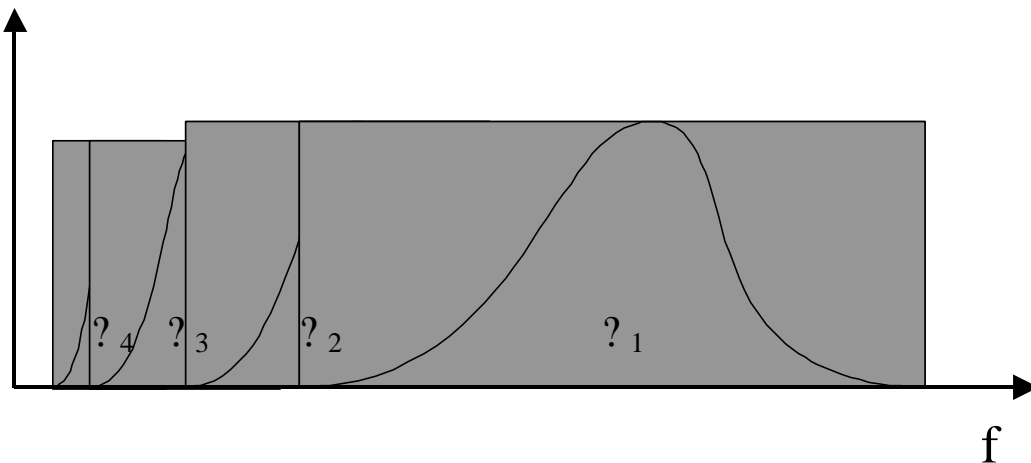


Figure 4. Wavelet Filter Bank

2.3.3. Subband coding and Multiresolution analysis

As we already stated that the wavelets can be used to implement a filter bank. The actual implementation is quite straightforward. The filtering is done iteratively so that the upper half of the spectrum is analyzed by the wavelet, and the lower part continues on to the next round.

Our wavelet itself acts as a half-band high-pass filter. As the result is bound to be redundant, we can discard every second value after the filtering by sub-sampling.

The lower half of the spectrum is filtered with the so-called scaling function. The wavelet and the scaling function form an orthonormal basis. Because the scaling function is a half-band low-pass filter, we get a similar effect as we did with the high-pass filter. The number of samples became twice the size of the spectrum. We can discard half of the samples simply by sub-sampling by two.

The result of the latter filtering becomes the data for the next round. This algorithm can be carried out until we are left with a single value for the low-pass spectrum.

The final result is a spectrum of equal size of that of the signal. The high-pass filterings are placed in the result vector so that the remaining single value (average) goes to the first “slot” in the vector. This algorithm is called sub-band coding. There exist also algorithms where the “zooming” algorithm is applied on both the low and the highpass components of each stage.

For some wavelet bases such as Daubechies wavelets, it is possible to construct a DWT equivalent for FFT that transforms a signal in linear time.

The following figure tries to visualize the subband-coding algorithm.

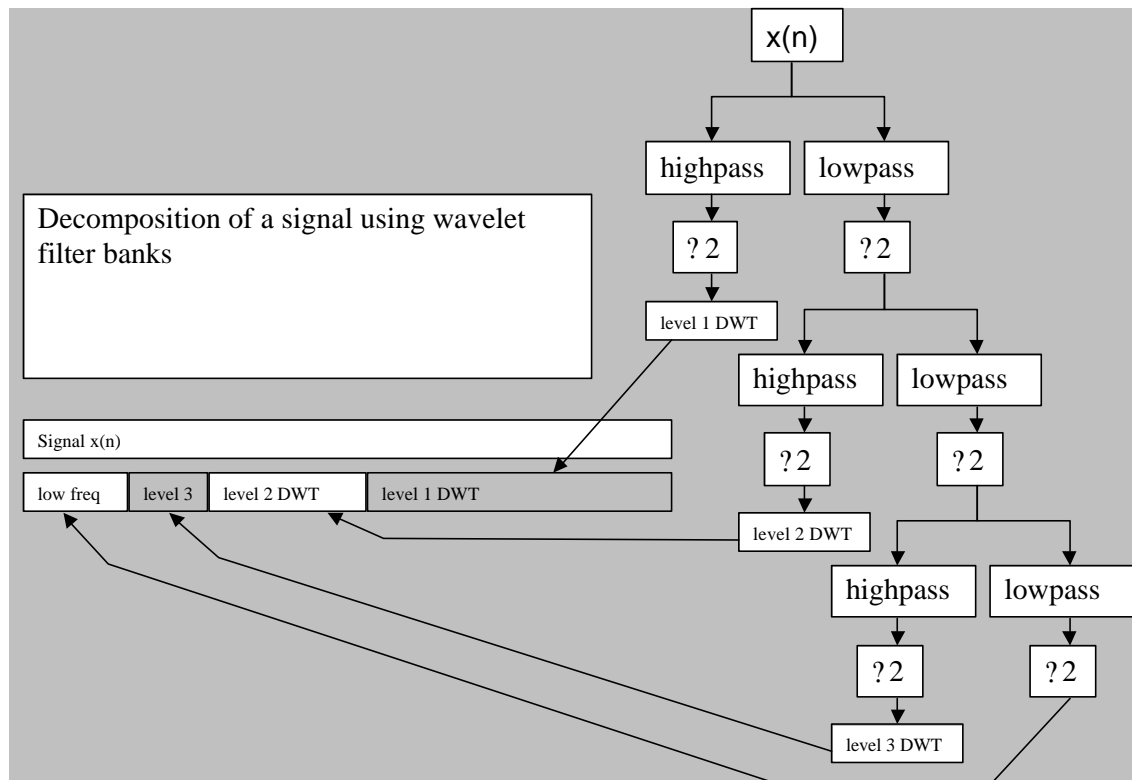


Figure 5. Multiresolution analysis – Subband coding scheme

3. References

- [1] C. Valens, "A Really Friendly Guide to Wavelets",
[<http://perso.wanadoo.fr/polyvalens/clemens/wavelets/wavelets.html>]
- [2] R. Polikar, "The Wavelet Tutorial",
[<http://engineering.rowan.edu/~polikar/WAVELETS/WTtutorial.html>]
- [3] N. Gershenfeld, "The Nature of mathematical Modeling" (Cambridge University Press, 1998)
- [4] A. Graps, "An Introduction to Wavelets" (IEEE Computational Science and Engineering 1995)