DCC-Lab 2020

TECHREP - Wavelet Transform VS Fourier Transform

Valérie Pineau Noël (valerie.pineau-noel.1@ulaval.ca)

This document gathers all information one needs to better understand the Fourier and the wavelet transforms. The first section will describe briefly the concept of Fourier transform : the general definition will be given for the classic continuous Fourier transform followed by the definitions of some related transforms. The second section will present the concept of wavelet transform. First, this section will describe the important characteristics of wavelets to better understand their use in comparision with the Fourier transform. Then, wavelets series and transforms will be defined.

Anyone with this link can edit the project : https://www.overleaf.com/5296336591bhhkdjgqfrby

Contents

1 Fourier Transform **3**
1.1 Discrete Fourier Transform . . . 5
1.2 Short-time Fourier Transform . . . 6

2 Wavelet Transform **7**
2.1 Important Characteristics of Wavelets . . . 7
2.2 Wavelet Expansion Series . . . 8
2.3 Classes of Wavelet Transform . . . 10
2.3.1 Discrete Wavelet Transform . . . 10
2.3.2 Continuous Wavelet Transform . . . 11

3 Fast Fourier and Wavelet Transforms **12**

Bibliography **13**

1 Fourier Transform

The Fourier transform is a mathematical tool used to represent a function in time or space by a sum of sines and cosines that contain the same amount of information as the original function[1]. It originates from the Fourier's Theorem, which states that *a function $f(x)$, having a period λ, can be synthesized by a sum of harmonic functions whose wavelengths are integral submultiples of* λ [2]. It is mainly used to decompose a signal into a sum of its frequency components, called the Fourier series. It is useful in many applications since some differential equations and convolutions are easier to solve in the frequency domain as opposed to the time domain.

Thus, any periodic function can be represented as a Fourier series :

$$f(x) = \frac{A_0}{2} + \sum_{m=1}^{\infty} A_m \cos(mkx) + \sum_{m=1}^{\infty} B_m \sin(mkx) \tag{1}$$

where $m \in \mathbb{N}$; A_0, A_m and B_m are the coefficients for the periodic function $f(x)$; $k = \frac{2\pi}{\lambda}$ and λ is the period (sometimes called the wavelength) of $f(x)$. The process of determining the values of the coefficients in this equation is referred to as Fourier analysis. Without going into details, the Fourier coefficients can be found using these equations :

$$A_0 = \frac{2}{\lambda} \int_0^{\lambda} f(x) dx \tag{2}$$

$$A_m = \frac{2}{\lambda} \int_0^{\lambda} f(x) \cos(mkx) dx \tag{3}$$

$$B_m = \frac{2}{\lambda} \int_0^{\lambda} f(x) \sin(mkx) dx \tag{4}$$

See [2] for more details.

The Fourier transform is known as a function $f(x)$ as follows :

$$F(k) = \int_{-\infty}^{+\infty} f(x) e^{ikx} dx \tag{5}$$

and the inverse Fourier transform is defined as follows :

$$f(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} F(x) e^{-ikx} dk \tag{6}$$

In this first definition (5), one can notice that the function $f(x)$ is multiplied by a complex number ; the modulus corresponds to the contribution of each frequency in the original function and the argument is the phase shift of their respective basic sinusoids. The function $F(k)$ is called the spectrum of $f(x)$. The variable k corresponds to the angular frequency $k = 2\pi v$ where v is the velocity in m^{-1} or s^{-1} since x can be in meters or in seconds according to its statement in space or time. As for the inverse Fourier transform, it converts data from the frequency domain to the time domain.

An example of a sum of two functions can be seen in Figure 1. The function represented in red can be a signal recorded from any device to which one would like to analyze its frequency composition.

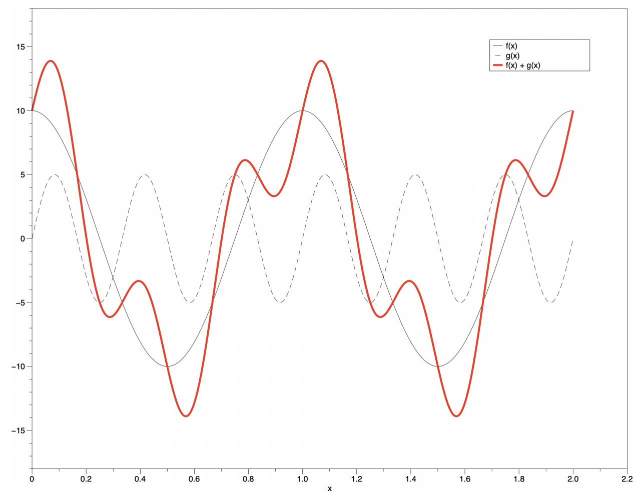

Figure 1: Sum of two functions, f(x) and g(x), shown in full ans dashed lines respectively.

Another way of presenting it could be using a frequency spectrum (Figure 2). Indeed it might be useful to have a reprensentation of every frequencies composing the original function according to their amplitude. In the case of figure 1, only two points are in the frequency spectrum since two waves are summed.

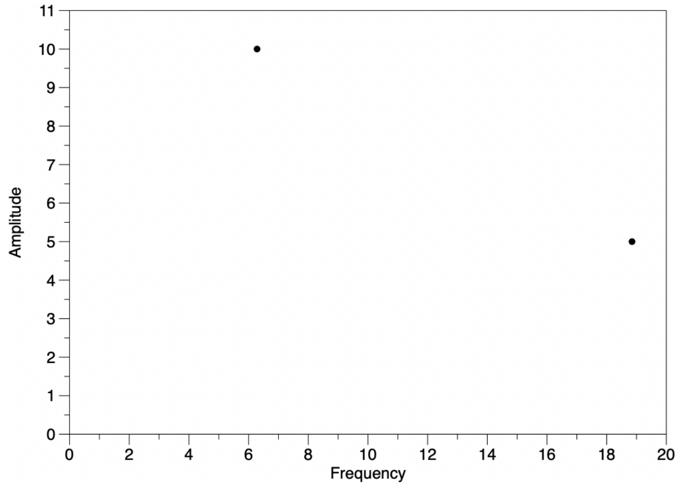

Figure 2: Frequency spectrum of the function f(x) + g(x) presented previously in Figure 1.

Notice that both mathematical definitions (5 and 6) do calculations for continuous functions. However, in the lab, measuring a signal is done by recording data at different time points ; a discret sampling is used instead of a contiuous sampling. The discrete Fourier transform is then applied in those cases.

1.1 Discrete Fourier Transform

The discrete Fourier transform is the equivalent of the Fourier transform, only that a finite number of points in a certain period of time are known from the original function [3]. In other words, the discrete Fourier transform is an estimate of the Fourier transform when only a limited collection of data are known [1, 2]. One should be careful when acquiring the data in an experiment, because the sampled points are considered to be typical of the whole function at all time.

The discrete Fourier transform can be defined as follows :

$$F[v_i] = \sum_{j=0}^{N-1} f[x_j] e^{i2\pi v_i x_j} \tag{7}$$

where $x_j = j\Delta$ corresponds to each measured point.

The inverse discrete Fourier transform is defined as follows :

$$f[x_i] = \frac{1}{N} \sum_{j=0}^{N-1} F[v_j] e^{-i2\pi v_j x_i} \tag{8}$$

Note that the discrete Fourier transform is defined with negative and positive frequencies, so from $-v_{max}$ to v_{max}. Also, F[0] corresponds to the mean of the signal.

Thus one could use the amplitude found with the transform to plot the distribution of energy, or power, of every component frequency. Indeed the amplitude of a harmonic wave is proportional to the amplitude squared. Since the discrete Fourier transform gives the amplitudes of all the components in the original function, the square of the transform can be used to determine the amount of energy at each component frequency. The graph obtained is called the *power spectrum*, which is function of spatial frequency. In other words, since the transform uses complex numbers, the power spectrum can be defined as the *product of the transform and its complex conjugate.* [2]

Sadly, this technique only works for periodic signal. Indeed, if the measured signal is non-periodic, sines and cosines, which are periodic functions, don't accurately represent the signal. Thus the analysis done in the frequency domain might be inaccurate. In those circumstances, the short-time Fourier transform can be used instead.

1.2 Short-time Fourier Transform

The short-time Fourier transform decomposes accurately a non-periodic function by using a time or space window that cuts this function into section to analyze their frequency content independently from one another. It's the same process than usually, but on many functions that together represent the main function. [3] In other words, it devides the function of interest into shorter segments of equal length. The Fourier transform is then computed on all of those segments individually. However, the function should be sectionned carefully since the lower frequency that the Fourier transform can catch is half of the length of the used window [4]. Indeed if a lower frequency is present in the original function, it won't be detected.

2 Wavelet Transform

Wavelets are commonly known as a *bref oscillation* with limited duration. Fourier and wavelet transforms are very similar : they are both linear operations, their inverse transformation matrices are the transpose of the original transformation matrices and they are localized in frequency. However, the decomposition of the signal does not produce a sum of sines and cosines in the wavelet transform ; it rather produces a sum of predefined wavelets.

2.1 Important Characteristics of Wavelets

One would use wavelets instead of Fourier transform if the function of interest has sharp discontinuities and/or spikes or if this function is non-periodic [3, 5]. Indeed wavelets enable the analysis to match with scale since they are non-local. Figure 3 shows four different families of wavelets. They have what is called a compact support, a defined scale in time, which allows the wavelet transform to make *multiresolution analysis* [6, 7]. As an example, the scale of wavelets in figure 3 are approximately 2000 for each family of wavelets presented here except for the Haar wavelet family that has a 500 scale. In other words, it just means that the wavelet function doesn't last forever. This is very different from the Fourier transform ; remember that a function in the Fourier domain in represented only with sines and cosines, functions that are streched out to infinity.

Thus, for wavelet transforms, *individual wavelet functions are localized in space and time* [3, 5]. Indeed, as explained previously and showed in figure 2, the distribution of frequencies present in a function can be pointed out into a frequency spectrum using the Fourier transform, but this gives no information on the timing of those frequencies ; the time resolution is lost [3, 4]. The wavelet transform gives back this time resolution by varying the size of the wavelet scale. This concept of *scale* is in fact very apparent to the concept of *window* in the short-time Fourier transform, but the window's size is not fixed. Indeed, to isolate short discontinuities, the scale needs to be short ; to isolate low frequencies, the scale needs to be big. To do that, the wavelet scale changes inversely proportional to the analyzed frequency in the wavelet transform [4]. Stretching and compressing the wavelet function is refered to *dilation* or *scaling* and corresponds to the physical notion of scale [5, 8]. In summary, the wavelet transform changes the scale according to the analysed frequency so every frequency component is studied with a *resolution matched to its scale* [3, 5–7].

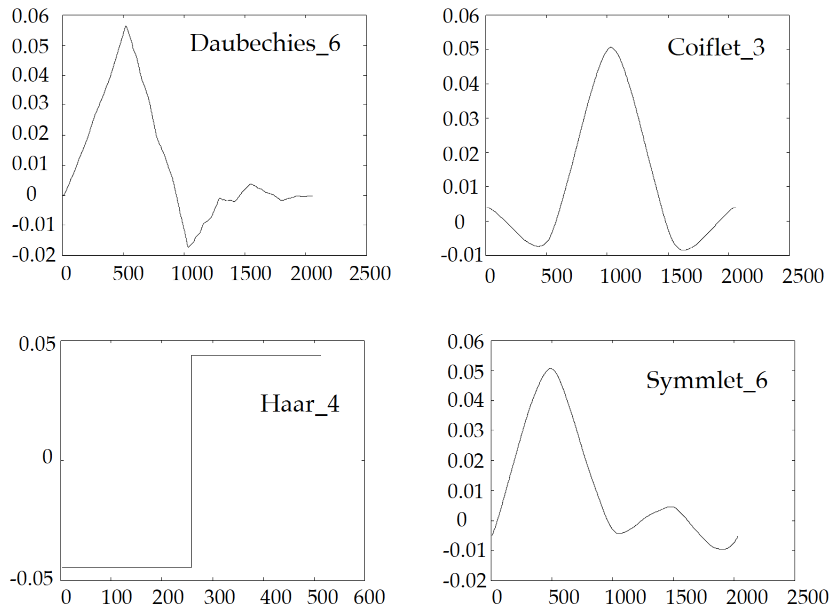

Figure 3: Four different families of wavelets. The number next to the family name represents the vanishing moments. (See Section 2.3.1) [3]

Then, the wavelet is translated on the original function to see where it matches to it. Thus the wavelet transform takes a one-dimensional function into a two-dimensional result because of the presence of the translation and scale parameters. Together, they define not only the frequency components of the function, but also the position of every frequency in the function [3, 5, 7].

Another important characteristic of the wavelet transform is that one can define the wavelet analyzing function $\Psi_{a,b}(t)$ with a specific family of wavelets to analyze its function. Indeed the wavelet transform does not have a single set of basis functions like the Fourier transform (sines and cosines). It can in fact uses different types of wavelets to extract information of the original function. [3, 4] Figure 3 shows some families of wavelets, but the number of possible kind of wavelets is infinite. The choice of the family of wavelets depends on the analyzed function and on the problem to be solved [6, 7].

2.2 Wavelet Expansion Series

Like equation 1 for the Fourier series, any function $f(t) \in L^2(R)$ can be represented as a linear combination of expansion functions :

$$f(t) = \sum_{b} c_{a0}(b)\varphi_{a0,b}(t) + \sum_{a=a0}^{\infty} d_a(b)\Psi_{a,b}(t) \tag{9}$$

where

- b (the integer translation parameter) and a (the scale parameter) $\in \mathbb{Z}$;
- c_{a0} : scaling coefficient (approximation) ;
- $\varphi_{a0,b}(t)$: scaling function ;
- d_a : wavelet coefficient (detail) ;
- $\Psi_{a,b}(t)$: wavelet analyzing function (sometimes called the mother function) ;
- $a0$: arbitrary starting scale.

If the expansion functions are orthogonal, the scaling coefficient c_{a0} and the wavelet coefficient d_a can be calculated using these equations :

$$c_{a0}(b) = \langle f(t), \varphi_{a0,b}(t) \rangle = \int f(t)\varphi_{a0,b}(t)dt \tag{10}$$

$$d_a(b) = \langle f(t)\Psi_{a,b}(t) \rangle = \int f(t)\Psi_{a,b}(t)dt \tag{11}$$

If the expansion functions are biorthogonal, $\varphi_{a0,b}(t)$ and $\Psi_{a,b}(t)$ must be replaced by their respective conjugate $\overline{\varphi_{a0,b}(t)}$ and $\overline{\Psi_{a,b}(t)}$ in both equations. See [7] for more details.

In the Fourier domain, sines and cosines are summed together in the Fourier series to represent a periodic function. For wavelet series, the first summation with the scaling function $\varphi_{a0,b}(t)$ is used to approximate the function $f(t)$ at the $a0$ scale. It's also considered as the average of the function $f(t)$. The rest of the summation at other scales is using the wavelet function $\Psi_{a,b}(t)$ which has a better resolution than the scaling function $\varphi_{a0,b}(t)$. It provides increasing details in the wavelet series. Indeed, as higher scales are added, the approximation made by the wavelet series becomes a more precise representation of the function $f(t)$ [7]. This is very similar to the Fourier series, where adding more and more terms in the summation increases the quality of the representation of the function of interest.

Now that the two terms of wavelet expansion series have been defined, the vanishing moment of wavelets can me introduced properly. Indeed, every familiy of wavelet have within itself subclasses of wavelets distinguished by an integer called the *vanishing moment m_k*.

$$m_k = \int_{-\infty}^{+\infty} f(t)t^k dx \tag{12}$$

where $k \in \mathbb{N}$. The moment k vanishes when $m_k = 0$. This integer is another important characteristic of wavelets since it is directly liked to the wavelet coefficient $d_a(b)$ [3]. A wavelet has m_k vanishing moments if and only if the scaling function $\varphi_{a0,b}(t)$ alone can generate polynomials up to degree m_k - 1. In other words, if a type of wavelet has three vanishing moments, polynomials up to the 3^{th} order $(g(x) = 1, x + 1, x^2 + 1)$ *will not* be identified by the wavelet function $\Psi_{a,b}(t)$; the wavelet coefficient $d_a(b) = 0$ for those conditions [9]. Considering this, the more vanishing moments a wavelet has, the more complex functions the scaling function $\varphi_{a0,b}(t)$ can represent [10] without the contribution of the wavelet analyzing function $\Psi_{a,b}(t)$. Ingrid Daubechies also proved that a wavelet with m_k vanishing moments must have a scale of at least $2m_k$ - 1 [5]. Thus the higher is the vanishing moment, the longer is the minimum support. In addition, the number of vanishing moments is linked to the wavelet oscillations, as it is shown in figure 4.

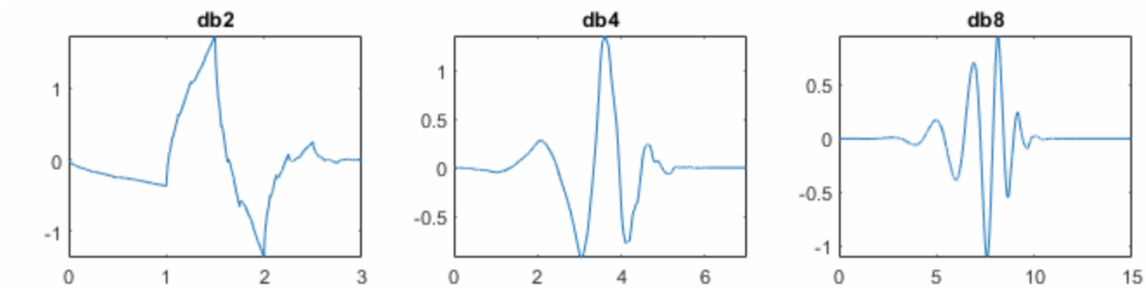

Figure 4: Daubechies wavelets with two, four and eight vanishing moments. (https://www.mathworks.com/help/wavelet/gs/choose-a-wavelet.html)

The vanishing moment is also an indicator of the regularity of the wavelet. Indeed a higher value of the wavelet vanishing moment m_k means a higher wavelet regularity. It also means that the better the wavelet regularity is, the smaller is the number of wavelet coefficients $d_a(b)$ used to represent the function of interest. It enables a smoother representation of the function $f(t)$ with sharper frequency resolution. Let's take the Haar wavelet as an example (Figure 3). Haar wavelets have bad regularity properties because they have abrupt changes at interval bounds, i.e. the scale changes too drastically. [6] Thus one would want a scaling function $\varphi_{a0.b(t)}$ with compact support as smooth as possible to have representative compact supports.

2.3 Classes of Wavelet Transform

The wavelet transforms are devided into two major classes according to their way of treating scale and translation. However, it's important to know that every family of wavelet has its own applications.

2.3.1 Discrete Wavelet Transform

If the function expanded is a sequence of points, like a discrete sampled signal, the resulting coefficients are called the discrete wavelet transform of $f(t)$. This transform restricts the values of scale and translation [4, 8]. Haar and Daubechies wavelets are examples of this class of wavelet transform.

The discrete wavelet transform is defined as follows :

$$W_{\varphi}(a0,b) = \frac{1}{\sqrt{M}} \sum_{x} f(t)\varphi_{a0,b}(t) \tag{13}$$

and

$$W_{\Psi}(a,b) = \frac{1}{\sqrt{M}} \sum_{x} f(t)\Psi_{a,b}(t) \tag{14}$$

so that

$$F(a,b) = W_{\varphi}(a0,b) + W_{\Psi}(a,b) \tag{15}$$

where M is the total number of known points in the original function $f(t)$. Depending on the value of the starting scale $a0$, the result of the discrete wavelet transform could change. Thus $W_{\varphi}(a0,b)$ and $W_{\Psi}(a,b)$ represent a family of transforms that differs in the starting scale $a0$ [7].

Discrete wavelet transforms are commonly used to denoise a certain signal. Indeed it identifies the noise part by determining the rapid changes in measurements, i.e. very high frequencies in the signal. The same method is used for image compression ; the algorithm finds the very high frequencies, which are noises, and eliminates it to get rid of some unnecessary information. [3, 4, 6, 7]

2.3.2 Continuous Wavelet Transform

In opposition to the discrete wavelet transform, the continuous wavelet transform can have its scale and translation set arbitrary [4, 8]. In other words, it's obtained by convolving a signal with an infinite number of functions with scale a and translation b using one wavelet analyzing function. This transform is the natural extension of the discrete wavelet transform [7]. The Morlet, Meyer and Mexican Hat wavelets are part of this class. The advantage of using the continuous wavelet transform is that the detection of features is easier, but it costs a lot more computation.

The definition of the continuous wavelet transform - and the general definition of the concept of wavelet transform - is as follows :

$$F(a,b) = \int_{-\infty}^{+\infty} f(t)\Psi_{a,b}(t)dt \tag{16}$$

Notice here that the continuous wavelet transform is represented in terms of wavelets alone. Because the starting scale $a0 = -\infty$, it eliminates the explicit scaling function dependance [7]. Thus the resulting continuous wavelet transform is a set of transform coefficients which measures the similarity of the function $f(t)$ with a set of basis functions $\Psi a,b(t)$ that is infinite.

3 Fast Fourier and Wavelet Transforms

When any transform, of any kind, is applied to a function, it has to identify every frequency present in it and their respective contribution. Computing the algorithm directly from the definition might be very long in some cases. Instead, one could use the fast version of the transform to reduce the complexity of the computation and fasten the procedure. Here, the fast Fourier transform will be shortly presented, but the same principle is applied for the fast wavelet transform. [3]

Normally, the discrete Fourier transform produces a matrix, which is commonly called a transformation matrix, that is multiplied to the signal to extract frequency components. A factor is usually multiplying the transformation matrix to keep the transformation unitary, i.e. to preserve the same energy in the physical and the Fourier domain. Also one should know that this transformation matrix is equal to the number of sample points n. Since the multiplication of an $n \times n$ matrix to a vector costs on the order of n^2 arithmetic operations, the computation becomes complicated very rapidly as the number of sampling points increases [3]. To decrease the computational cost, the fast Fourier transform factorizes the transformation matrix into a product of multiple sparse matrices, which are mostly zeros. Consequently, this manipulation decreases the complexity of the algorithm and computes the discrete Fourier transform faster. [2, 7]

References

[1] D. C. Côté, "La transformée de fourier - un guide pratique", Cours d'Optique et optique instrumentale (GPH-2005), Department of physics, physics engineering and optics, Laval University, Québec, Canada (2014).

[2] E. Hecht, "Optics - 4th edition", Pearson Education **0-8053-8566-5** (2002).

[3] A. Graps, "Van introduction to wavelets", Institute of Electrical and Electronics Engineers Society **2** (1995).

[4] M. Ryan, *What is wavelet and how we use it for data science*, (2019) `https://towardsdatascience.com/what-is-wavelet-and-how-we-use-it-for-data-science-d19427699cef` (visited on 03/17/2020).

[5] I. Daubechies, "Ten lectures on wavelets", Society for Industrial and Applied Mathematics (1992).

[6] I. Dremin, "Wavelets and their use", Lebedev Physical Institute **[arXiv:hep-ph/0101182]** (2001).

[7] R. C. Gonzalez and R. E. Woods, "Digital image processing - 2nd edition", Springer (2001).

[8] T. M. Inc., *Continuous wavelet transform and scale-based analysis*, (2020) `https://www.mathworks.com/help/wavelet/gs/continuous-wavelet-transform-and-scale-based-analysis.html` (visited on 03/19/2020).

[9] Z. P. G. Wang, "Study on optimal selection of wavelet vanishing moments for ecg denoising", Scientific Reports **7** (2017).

[10] L. Chun-Lin, "A tutorial of the wavelet transform", (2010).