Chapter 5

Wavelets and Fourier transforms

5.1 Introduction

In this chapter we continue studying transformation methods. The main idea behind the techniques presented is to transform data from the spatial domain to a frequency domain. We start by introducing the Fourier transform since it is simpler and historically older.

5.2 Dynamic Fourier Analysis

A physical process can be described either in the time domain, by the values of some quantity f as a function of time t, for example, $f[t]$, or in the frequency domain, where the process is specified by giving its amplitude H (generally a complex number) as a function of frequency n, that is $F[n]$, with $-\infty < n < \infty$. For many applications, it is useful to think of $f[t]$ and $F[n]$ as being two different representations of the same function. Generally, these two representations are connected using Fourier transform formulas:

$$F[n] = \int_{-\infty}^{\infty} f[t] e^{2\pi i n t} dt = \int_{-\infty}^{\infty} f[t] \left(\cos(2\pi n t) + i \sin(2\pi n t)\right) dt \qquad (5.1)$$
$$f[t] = \int_{-\infty}^{\infty} F[n] e^{-2\pi i n t} dn,$$

where, $F[n]$ represents the Fourier transform and $f[t]$ represents the inverse Fourier transform. The Fourier transform $F[n]$ converts the data from the time domain into the frequency domain. The inverse Fourier transform $f[t]$ converts the frequency domain components back into the original time-domain

signal. A frequency-domain plot shows how much of the signal lies within each given frequency band over a range of frequencies.

In order to analyze a statistical time series using Fourier transforms we need to assume that the structure of the statistical or stochastic process which generates the observations is essentially invariant through time. This assumption is summarized in the condition of stationarity which states that its finite dimensional distribution remains the same throughout time.

This condition is hard to verify and we generally enforce a weak stationarity condition. For example, a time series x_t is weak stationary, if its second-order behavior remains the same, regardless of the time t. Looking at the Fourier transform representation (5.1) we see that a stationary series is represented as a superposition of sines and cosines that oscillate at various frequencies. Therefore, a stationary time series can be matched with its sine and cosine series representation (these are the representative frequencies).

Next we introduce some definitions that will be used when applying the Dynamic Fourier Analysis.

5.2.1 Tapering

Generally, calculating the continuous Fourier transform and its inverse (5.1) is a hard problem in practice. For this reason, a Discrete Fourier Transform (DFT) is used when we observe discrete time series with a finite number of samples from a process that is continuous in time. We will discuss the DFT in subsection 5.2.3 however applying it requires an extra step known as tapering.

When the original function (process) is discontinuous the corresponding signal value abruptly jumps, yielding spectral leakage (that is, the input signal does not complete a whole number of cycles within the DFT time window). In this case to perform DFT we need to multiply the finite sampled times series by a windowing function, or "a taper". The taper is a function that smoothly decays to zero near the ends of each window and it is aimed at minimizing the effect of the discontinuity by decreasing the time series magnitude so it approaches zero at the ends of the window. Although, spectral leakage cannot be prevented it can be significantly reduced by changing the shape of the taper function in a way to minimize strong discontinuities close to the window edges. In seismic data analysis cosine tapers are often used since it is both effective and easy to calculate. The cosine taper can be written as:

$$c(t) = \begin{cases} \frac{1}{2}\left(1 - \cos\frac{\pi t}{a}\right) & 0 \leq t \leq a \\ 1 & a \leq t \leq (1-a) \\ \frac{1}{2}\left(1 - \cos\frac{\pi}{a}(1-t)\right) & (1-a) \leq t \leq 1 \end{cases}$$

where t is time and a is the taper ratio. The cosine window represents an attempt to smoothly set the data to zero at the boundaries while not significantly reducing the level of the values within the window. This form of tapering reduces the leakage of the spectral power from a spectral peak to frequencies

far away and it coarsens the spectral resolution by a factor $1/(1-a)$ for the above cosine tapers.

In general, the effects of applying a taper are:

1. Decrease the time series magnitude to zero or near zero at its start and end so that there is no sharp discontinuity between the first and last point in the periodic time series.

2. Change the weighting of samples in the time series so that those near the middle contribute more to the Fourier transform.

3. Reduce the resolution of the spectrum by averaging adjacent samples.

For non-stationary time series signals, tapering may bias the spectral amplitudes even if the taper is normalized. However, for these we should not use Fourier transform in the first place.

Generally, it is difficult to give precise recommendations on which tapers to use in all specific situation. The work of [24] recommends to reduce the leakage by increasing the length N of the time window and at the same time decreasing the taper ratio a, such that the length (Na) of the cosine half-bells is kept constant.

In practice, a value of $a = 5\%$ for window lengths of 30 seconds or 60 seconds is good enough for frequencies down to 0.2 Hz. See [165] for more details.

5.2.2 Estimation of spectral density with Daniell Kernel

In this section we present a probabilistic approach to estimate the magnitudes and frequency coefficients in a Fourier Transform. A stationary process X_t may be defined by taking linear combinations of the form

$$X_t = \sum_{j=1}^{m} (A_j cos(2\pi\lambda_j t) + B_j sin(2\pi\lambda_j t)) \tag{5.2}$$

where $0 \leq \lambda \leq \frac{1}{2}$ is a fixed constant and $A_1, B_1, A_2, B_2, ...A_m, B_m$ are all uncorrelated random variables with mean zero and

$$Var(A_j) = Var(B_j) = \sigma_j^2.$$

We assume $\sum_{j=1}^{m} \sigma_j^2 = \sigma^2$ so that the variance of the process X_t is σ^2, and let the spectral density $f(\lambda)$ satisfy the equation

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} f(\lambda) d\lambda = \sigma^2.$$

Then, the process given in (5.2) converges to a stationary process with spectral density f as $m \to \infty$.

In order to estimate the spectral density f, we define the estimators as a weighted average of periodogram values (I) for frequencies in the range $(j-m)/n$ to $(j+m)/n$. In particular we define:

$$\hat{f}(j/n) = \sum_{k=-m}^{m} W_m(k) I\left(\frac{j+k}{n}\right)$$

The set of weights $\{W_m(k)\}$ sum to one and the set is often referred to as a kernel or a spectral window. Essentially, this kernel with parameter m is a centered moving average which creates a smoothed value at time t by averaging all values between and including times $t-m$ and $t+m$.

We define:

$$W_m(k) = \frac{1}{2m+1} \quad \text{for } -m \leq k \leq m, \ \sum_k W_m(k) = 1 \quad \text{and } \sum_k k W_m(k) = 0$$

The smoothing formula $\{u_t\}$ for a Daniell kernel with $m = 1$ corresponds to the three weights $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ is given by:

$$\hat{u}_t = \frac{1}{3}(u_{t-1} + u_t + u_{t+1})$$

Applying the Daniell kernel again on smoothed values $\{\hat{u}_t\}$ produces a more extensive smoothing by averaging over a wider time interval.

$$\hat{\hat{u}}_t = \frac{\hat{u}_{t-1} + \hat{u}_t + \hat{u}_{t+1}}{3} = \frac{1}{9}u_{t-2} + \frac{1}{9}u_{t-2} + \frac{3}{9}u_t + \frac{2}{9}u_{t+1} + \frac{1}{9}u_{t+2} \qquad (5.3)$$

Consequently applying the Daniell kernel transforms the spectral windows into a form of Gaussian probability density function.

5.2.3 Discrete Fourier Transform

The Discrete Fourier Transform (DFT) is the equivalent of the continuous Fourier transform for signals known only at N instants separated by sample times T (i.e. a finite sequence of data). It is commonly used in practical applications since we generally cannot observe a signal continuously. The (DFT) converts a finite sequence of equally-spaced samples of a function into a sequence of equally-spaced samples of the discrete-time Fourier transform (DTFT), which is a complex-valued function of the frequency.

The DFT is the most important discrete transform, used to perform Fourier analysis in many practical applications. In image processing, the samples can be the values of pixels along a row or column of an image. The DFT can also be used to efficiently solve partial differential equations, and to perform other operations such as convolutions or multiplying large integers.

The discrete Fourier transform of the sequence $f[k]$ is defined as:

$$F[n] = \sum_{k=0}^{N-1} f[k] e^{-j\frac{2\pi}{N}nk}, n = 0, \ldots, N-1 \tag{5.4}$$

We observe that equation (5.4) can be written in matrix form as:

$$\begin{bmatrix} F[0] \\ F[1] \\ F[2] \\ F[3] \\ \vdots \\ F[N-1] \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & \ldots & 1 \\ 1 & W & W^2 & W^3 & \ldots & W^{N-1} \\ 1 & W^2 & W^4 & W^6 & \ldots & W^{N-2} \\ 1 & W^3 & W^6 & W^9 & \ldots & W^{N-3} \\ \vdots & & & & & \\ 1 & W^{N-1} & W^{N-2} & W^{N-3} & \ldots & W \end{bmatrix} \begin{bmatrix} f[0] \\ f[1] \\ f[2] \\ f[3] \\ \vdots \\ f[N-1] \end{bmatrix}$$

where $W = \exp(-j2\pi/N)$. Since $W^N = 1$ note that the columns are powers of the initial column. In linear algebra this is a particular type of a Vandermonde matrix and multiplying vectors with it is a very fast operation.

The inverse discrete Fourier transform corresponding to (5.4) is

$$f[k] = \frac{1}{N} \sum_{=0}^{N-1} F[n] e^{j\frac{2\pi}{N}nk}, k = 0, \ldots, N-1 \tag{5.5}$$

Thus the inverse matrix is $1/N$ times the complex conjugate of the original matrix.

In the process of taking the inverse transform the terms $F[n]$ and $F[N-n]$ combine to produce 2 frequency components, only one of which is considered to be valid (the one at the lower of the two frequencies). From (5.5), the contribution to $f[k]$ of $F[n]$ and $F[N-n]$ is:

$$f_n[k] = \frac{1}{N}\{F[n]e^{j\frac{2\pi}{N}nk} + F[N-n]e^{j\frac{2\pi}{N-n}k}\} \tag{5.6}$$

For all $f[k]$ real, $F[N-n] = \sum_{k=0}^{N-1} f[k] e^{-j\frac{2\pi}{N}(N-n)k}$.

The time taken to evaluate a DFT on a digital computer depends principally on the number of multiplications involved, since these are the slowest operations. With the DFT, this number is directly related to N^2 (matrix multiplication of a vector), where N is the length of the transform. For most problems, N is chosen to be at least 256 in order to get a reasonable approximation for the spectrum of the sequence under consideration, hence computational speed becomes a major consideration. Efficient computer algorithms for estimating discrete Fourier transforms are known as Fast Fourier Transforms (FFT) algorithms (subsection 5.2.4). These algorithms rely on the fact that the standard DFT involves a lot of redundant calculations.

For instance, we can rewrite

$$F[n] = \sum_{k=0}^{N-1} f[k] e^{-j\frac{2\pi}{N}nk}, n = 0, \ldots, N-1$$

as

$$F[n] = \sum_{k=0}^{N-1} f[k] W_N^{nk}, n = 0, \ldots, N-1 \tag{5.7}$$

We observe that in (5.7), the same values of W_N^{nk} are calculated many times as the computation proceeds.

5.2.4 The Fast Fourier Transform (FFT) Method

The work of Heston [95] in stochastic volatility modeling, marks the first time Fourier inversion appears in option pricing literature. Since then, Fourier inversion has become more and more prevalent. This is due to the fact that analytic formulas are hard to calculate and can only be done in the simplest of models.

The fast Fourier transform (FFT) allows us to compute efficiently sums of the form:

$$W_m = \sum_{j=1}^{N} \exp\left(\frac{2\pi i(j-1)(m-1)}{N}\right) x_k, \quad [28] \tag{5.8}$$

with a complexity of $O(N \log N)$. The big O notation is a mathematical notation that describes the limiting behavior of a function when the argument tends towards a particular value. It will be described in details in chapter 6.

Discretizing the semi-definite integral with equidistant quadrature roots $\Delta x = x_{i+1} - x_i$ gives:

$$\int_0^\infty e^{-ivk} \psi_\alpha(v) dv \approx \sum_{j=1}^{N} e^{-i(j-1)\Delta x k} \psi_\alpha(x_j). \tag{5.9}$$

In finance this expression will allow us to simultaneously compute option prices for a range of N strikes. We choose

$$k = -\frac{N\Delta k}{2} + (m-1)\Delta k, \quad m = 1, \ldots, N. \tag{5.10}$$

Here the discretization in (log)-strike domain and the quadrature discretization have to obey the Nyquist relation:

$$\Delta k \Delta x = \frac{2\pi}{N}, \quad [63] \tag{5.11}$$

which effectively restricts the choice of the integration domain and the range of strikes to be computed. Therefore, in practice we have to balance the interpolation error in the strike domain against the quadrature error. The quadrature error comes from the numerical integration.

The Fourier Transform methods arise naturally with the Lévy Khintchine formula (see [120] and [21]). In fact, the characteristic function of a distribution μ corresponds to the Fourier transform of the density function. The

characteristic function of a distribution μ on $\mathbb{R}$ is defined as

$$\phi_\mu(u) = \hat{\mu}(u) = \int_{\mathbb{R}} e^{iux}\mu(dx). \tag{5.12}$$

In the case where μ is a probability density function, its Fourier Transform may be written in term of its moments. If we define the n-th moment of μ by $\mu_n = \int x^n \mu(dx)$, then from (5.12) we obtain

$$\phi_\mu(u) = \int_{\mathbb{R}} \sum_{n=0}^{\infty} \frac{(iux)^n}{n!}\mu(dx) = \sum_{n=0}^{\infty} \frac{(iu)^n}{n!}\mu_n.$$

Fourier transform methods have become popular in finance in the context of exponential Lévy processes primarily due to lack of knowledge of the density functions in closed form and the impossibility to obtain closed form expressions for option prices. We will study Lévy processes in chapter 12 of this book.

As we stated, the fast Fourier transform methods have become a very important tool in finance. The fact that many characteristic functions can be expressed analytically using it, makes the Fourier transform a natural approach to compute option prices.

Theorem 5.2.1. *Let f be a function $f : \mathbb{R} \to \mathbb{R}$, continuous, bounded in $L^1(\mathbb{R})$, and so that $F(f) \in L^1(\mathbb{R})$. Then*

$$E[f(X_t)] = F^{-1}[\tilde{f}](.)\phi_{X_t}(.)]_0,$$

with F denoting the Fourier transform operator, and $\tilde{f}(x) = f(-x)$.

Proof. We observe that

$$\int_{\mathbb{R}} f(x)\mu(dx) = \int_{\mathbb{R}} f(y-x)\mu(dx)|_{y=0},$$

and the convolution product is integrable since f is bounded and continuous. Hence the theorem follows. □

This theorem allows us to compute option values which are typically expressed as expectations of the final payoff f from the Fourier transform of the density of X_t (its characteristic function). When applied to the option pricing problem, the assumptions on f are usually not satisfied since most payoff functions are not in $L^1(\mathbb{R})$. The payoff function is usually replaced with a truncated version to satisfy the assumption.

Two Fourier transformations are possible, one with respect to the log strike price of the option, the other with respect to the log spot price at T. We will consider the first approach.

Theorem 5.2.2. *Define the truncated time value option by*

$$z_T(k) = E[(e^{X_T} - e^k)_+] - (1 - e^k)_+,$$

then for an exponential martingale e^{X_T} such that $E[e^{(1+\alpha)X_T}] < \infty$, the Fourier transform of the time value of the option is defined by

$$\zeta_T(v) = \int_{-\infty}^{\infty} e^{ivk} z_T(k)dk = e^{ivrT}\frac{\Phi_T(v-i)-1}{iv(1+iv)}.$$

The price of the option is then given by

$$C(t,x) = x[(1-e^{k_t}) + \frac{1}{2\pi}\int_{\mathbb{R}} \zeta_T(v)e^{-ivk_t}dv].$$

Proof. We sketch an outline of the proof: The discounted stock process $\tilde{S}_t = e^{X_t}$ is a martingale. Therefore

$$\int_{-\infty}^{\infty} e^x q_{X_t}(x)dx = 1.$$

The stock process admits a moment of order $1+\alpha$ so that there exists $\alpha > 0$, such that

$$\int_{-\infty}^{\infty} e^{(1+\alpha)x} q_{X_t}(x)dx < 0,$$

which can be used to show both integrals $\int_{-\infty}^{\infty} xe^x q_{X_t}(x)$ and $\int_{-\infty}^{\infty} e^x q_{X_t}(x)dx$ are finite. Using Fubini's theorem, we explicitly compute the Fourier transform z_T as follows: We write

$$\begin{aligned} z_T(k) &= E[(e^{X_T} - e^k)_+] - (1-e^k)_+ \\ &= \int_{-\infty}^{\infty} [(e^x - e^k)1_{x\geq k} - (1-e^k)1_{0\geq k}]q_{X_t}(x)dx \\ &\int_{-\infty}^{\infty} [(e^x - e^k)(1_{x\geq k} - 1_{0\geq k})]q_{X_t}(x)dx, \end{aligned}$$

where in the second term we use the fact that $\tilde{S}_t$ is a martingale. Assuming we can interchange the integrals, the Fourier transform of z_T is given by

$$\begin{aligned} F(z_T)[\nu] &= \int_{-\infty}^{\infty} e^{i\nu k} z_T(k)dk \\ &= \int_{-\infty}^{\infty} e^{i\nu k} \int_{-\infty}^{\infty} (e^x - e^k)(1_{x\geq k} - 1_{0\geq k})q_{X_t}(x)dxdk \\ &= \int_{-\infty}^{\infty} q_{X_t}(x) \int_{-\infty}^{\infty} e^{i\nu k}(e^x - e^k)(1_{x\geq k} - 1_{0\geq k})dkdx \\ &= \int_{-\infty}^{\infty} q_{X_t}(x) \int_{0}^{x} e^{i\nu k}(e^x - e^k)dkdx. \end{aligned}$$

When applying the Fubini's theorem, we observe that when $x \geq 0$, $\int_0^x |e^{i\nu k}(e^x - e^k)| \leq e^x(x-1)dk$, and $\int_{-\infty}^{\infty} e^x(x-1)q_{X_t}(x)dx < \infty$. Also, when $x < 0$,

$\int_0^x |e^{iνk}(e^x - e^k)| \leq (e^x - 1)dk$, and $\int_{-\infty}^{\infty}(e^x - 1)q_{X_t}(x)dx < \infty$. Thus,

$$\begin{aligned} F(z_t)[\nu] &= \int_{-\infty}^{\infty} q_{X_t}(x) \int_0^x e^{i\nu k}(e^x - e^k)dkdx \\ &= \int_{-\infty}^{\infty} q_{X_t}(x) \left[\frac{e^x}{i\nu}(e^{i\nu x} - 1) - \frac{1}{i\nu + 1}(e^{(i\nu+1)x} - 1)\right] dx \\ &= \int_{-\infty}^{\infty} q_{X_t}(x) \frac{e^{(i\nu+1)x} - 1}{i\nu(i\nu + 1)} dx. \end{aligned}$$

□

This approach has the advantage of not requiring an exact value of α in order to compute the option prices. The numerical implementation for the use of this truncation function, which is not smooth, has the inconvenience of generating rather large truncation errors, making the convergence slow. Thus we may observe that the Fourier transform of the time value of the option is

$$\zeta_T(v) = e^{ivrT} \frac{\Phi_T(v - i) - 1}{iv(1 + iv)} \sim |v|^{-2}.$$

5.3 Wavelets theory

Recall that in order to perform Fourier decomposition the time series as well as the process needs to be stationary. From a regression point of view, we may imagine a system responding to various driving frequencies by producing linear combinations of sine and cosine functions. If a time series X_t is stationary, its mean, variance and covariances remains the same regardless of the time t, therefore we can match a stationary time series with sines and cosines because they behave the same indefinitely.

Non-stationary time series require a deeper analysis. The concept of wavelet analysis generalize dynamic Fourier analysis, with functions that are better suited to capture the local behavior of nonstationary time series. These functions are called wavelet functions.

Wavelet transform is often compared with the Fourier transform, in the sense that we have a similar decomposition with wavelets replacing the sene functions. In fact, the Fourier transform may be viewed as a special case of a continuous wavelet transform with the choice of the wavelet function

$$\psi(t) = e^{-2\pi it}\psi(t) = e^{-2\pi it}.$$

The main difference between a wavelet transform and the Fourier transform is the fact that the wavelets analysis is a more general method which is localized in both time and frequency and has a faster computation speed, whereas the standard Fourier analysis is only localized in frequency [188].

As a mathematical tool, wavelets can be used to extract information from various kinds of data, including digital signals, images and several others. Wavelets are well-suited for approximating data with sharp discontinuities ([83]).

5.3.1 Definition

The wavelet transform of a function $f(t)$ with finite energy is defined as the integral transform for a family of functions

$$\psi_{\lambda,t}(u) \equiv \frac{1}{\sqrt{\lambda}}\psi\left(\frac{u-t}{\lambda}\right)$$

and is given as

$$Wf(\lambda,t) = \int_{-\infty}^{\infty} f(u)\psi_{\lambda,t}(u)du \quad \lambda > 0. \tag{5.13}$$

The function ψ is called the wavelet function (mother wavelet). The λ is a scale parameter, t a location parameter and the generated functions $\psi_{\lambda,t}(u)$ are called wavelets. In the case where $\psi_{\lambda,t}(u)$ is complex, we use complex conjugate $\overline{\psi}_{\lambda,t}(u)$ in the definition (5.13). The normalizing constant $\frac{1}{\sqrt{\lambda}}$ is chosen so that

$$|\psi_{\lambda,t}(u)|^2 \equiv \int |\psi_{\lambda,t}(t)|^2 du = \int |\psi(t)|^2 dt = 1$$

for all the scales λ.

The choice of the mother wavelet $\psi(t)$ is critical. Different choices lead to different decompositions and the right choice very much depends on the particular data studied. The only property the mother wavelet function $\psi(t)$ has is unit energy (or variance in probability, i.e., $|\psi(t)|^2_{L^2} = \int |\psi(t)|^2 dt = 1$). However, to produce a good decomposition it must possess the following properties:

- fast decay as $|t| \to \infty$ in order to obtain localization in space;
- zero mean, that is, $\int \psi(t)dt = 0$.

These two properties above typically make the function look like a wave reflected around 0 which is why the function $\psi(t)$ is called a wavelet. Please refer to [75] for more details.

5.3.2 Wavelets and Time series

The wavelet function ψ is in effect a band-pass filter. When we scale it using λ for each level the bandwidth is reduced. This creates the problem that in order to cover the entire spectrum, an infinite number of levels would be required. To deal with this in practice a extra decomposition is used. Wavelets are generated by a scaling function (father wavelet), ϕ in addition to a mother wavelet function, ψ. The scaling function is used to capture the smooth, low-frequency nature of the data, whereas the wavelet functions are used to capture the detailed, and high-frequency nature of the data.

The scaling function integrates to one, and the wavelet function integrates to zero:

$$\int \phi(t)dt = 1 \quad \text{and} \quad \int \psi(t)dt = 0. \tag{5.14}$$

Generally, the analytic form of wavelets does not exist and they are typically generated using numerical schemes. Unlike in the Fourier transform case, here we talk about time and scale, rather than time and frequency. The departure from the periodic functions (sin and cos) means that frequency looses its precise meaning.

The orthogonal wavelet decomposition of a time series, x_t, $t = 1, \dots n$ is defined as,

$$\begin{aligned} x_t = &\sum_k s_{J,k}\phi_{J,k}(t) + \sum_k d_{J,k}\psi_{J,k}(t) \\ &+ \sum_k d_{J-1,k}\psi_{J-1,k}(t) + \ldots + \sum_k d_{1,k}\psi_{1,k}(t) \end{aligned} \tag{5.15}$$

where the $J-$th level represents the number of scales (frequencies) for the orthogonal wavelet decomposition of the time series x_t. The index k ranges from 1 to the number of coefficients associated with the specified component. The functions $\phi_{J,k}(t), \psi_{J,k}(t)$, $\psi_{J-1,k}(t), \dots, \psi_{1,k}(t)$ are generated from the scaling function, $\phi(t)$, and the wavelet function, $\psi(t)$, by shifting and scaling based on powers of 2:

$$\begin{aligned} \phi_{J,k}(t) &= 2^{\frac{-J}{2}}\phi\left(\frac{t-2^Jk}{2^J}\right), \\ \psi_{j,k}(t) &= 2^{\frac{-j}{2}}\psi\left(\frac{t-2^jk}{2^j}\right), \quad j = 1, \dots, J, \end{aligned}$$

where 2^jk is the shift parameter and 2^j is the scale parameter. The wavelet functions are spread out and shorter for larger values of j and narrow and taller for smaller values of the scale ([188]). The reciprocal of the scale parameter ($\frac{1}{2^j}$) in wavelet analysis is analogous to frequency ($\omega_j = \frac{j}{n}$) in Fourier analysis. This is as a result of the fact that larger values of the scale refer to slower, smoother movements of the signal, and smaller values of the scale refer to faster, finer movements of the signal.

The **discrete wavelet transform (DWT)** of a time series data x_t, is given by the coefficients $s_{J,k}$ and $d_{J-1,k}$ for $j = J, J-1, \dots, 1$ in (5.15). To some degree of approximation, they are given by

$$s_{J,k} = n^{-1/2}\sum_{t=1}^{n} x_t\phi_{J,k}(t), \tag{5.16}$$

$$d_{j,k} = n^{-1/2}\sum_{t=1}^{n} x_t\psi_{j,k}(t) \quad j = J, J-1, \dots, 1. \tag{5.17}$$

The magnitudes of these coefficients measure the importance of the corresponding wavelet term in describing the behavior of the time series x_t. The $s_{J,k}$ are known as the smooth coefficients because they describe the smooth behavior of the data. The $d_{j,k}$ are known as the detail coefficients since they represent the finer and high frequency nature of the time series data.

A way to measure the importance of each level is to evaluate the proportion of the total power or energy explained by the phenomenon under study. The level corresponds to the number of scales (frequencies) for the orthogonal wavelet decomposition of the time series. For example, in seismic studies larger values of the scale correspond to slower and smoother movements of the seismogram whereas, smaller values of the scale correspond to faster and finer movements of the seismogram.

The total energy, P of a time series x_t, for $t = 1, \ldots, n$, is

$$P = \sum_{t=1}^{n} x_t^2. \tag{5.18}$$

The total energy associated to each level of scale is

$$P_J = \sum_{k=1}^{n/2^J} s_{J,k}^2 \tag{5.19}$$

and

$$P_j = \sum_{k=1}^{n/2^j} d_{j,k}^2, \quad j = J, J-1, \ldots, 1. \tag{5.20}$$

Thus, we can rewrite equation (5.18) as

$$P = P_J + \sum_{j=1}^{J} P_j. \tag{5.21}$$

The proportion of the total energy explained by each level is the ratios of the total energy associated with each coefficient of detail to the total energy of the time series, or

$$P_j / P$$

for $j = J, J-1, \ldots, 1$. Please refer to [18] for more details.

5.4 Examples of Discrete wavelets transforms (DWT)

A discrete wavelet transform (DWT) is defined as any wavelet transform for which the wavelets are discretely sampled. Equations (5.16) and (5.17) exemplify such DWT. A discrete wavelet transform decomposes a signal into a set of mutually orthogonal wavelet basis functions. One important feature of the DWT over other transforms, such as the Fourier transform, lies in its ability to offer temporal resolution, i.e. it captures both frequency and location (or time) information.

5.4.1 Haar wavelets

The Haar DWT is one of the simplest possible wavelet transforms. One disadvantage of the Haar wavelet is the fact that it is not continuous, and therefore not differentiable. This property can, however, be an advantage for the analysis of signals with sudden transitions, such as monitoring of tool failure in machines (see e.g., [130]).

The Haar wavelet function $\psi(t)$ can be described as

$$\psi(t) = \begin{cases} 1, & 0 \leq t < \frac{1}{2} \\ -1, & \frac{1}{2} \leq t < 1 \\ 0, & \text{Otherwise.} \end{cases}$$

and the scaling function $\phi(t)$ can be described as,

$$\phi(t) = \begin{cases} 1, & 0 \leq t < 1 \\ 0, & \text{Otherwise.} \end{cases}$$

The Haar functions in fact are very useful for demonstrating the basic characteristics of wavelets (see [188]).

5.4.1.1 Haar functions

The family of N Haar functions $h_k(t), (k = 0, \ldots, N-1)$ are defined on the interval $0 \leq t \leq 1$. The shape of the specific function $h_k(t)$ of a given index k depends on two parameters p and q:

$$k = 2^p + q - 1 \tag{5.22}$$

For any value of $k \geq 0$, p and q are uniquely determined so that 2^p is the largest power of 2 contained in k ($2^p < k$) and $q - 1$ is the remainder $q - 1 = k - 2^p$.

Example 5.4.1. *When $N = 16$, the index k with the corresponding p and q are shown in Table* (5.4.1)*:*

Table 5.1: Determining p and q for $N = 16$

k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
p	0	0	1	1	2	2	2	3	3	3	3	3	3	3	3	3
q	0	1	1	2	1	2	3	4	1	2	3	4	5	6	7	8

Thus the Haar functions can be defined recursively as:

1. When $k = 0$, the Haar function is defined as a constant

$$h_0(t) = 1/\sqrt{N}$$

2. When $k > 0$, the Haar function is defined by

$$\psi(t) = \frac{1}{\sqrt{N}} \begin{cases} 2^{p/2}, & (q-1)/2^p \leq t < (q-0.5)/2^p \\ -2^{p/2}, & (q-0.5)/2^p \leq < t < (q-0.5)/2^p \\ 0, & \text{Otherwise.} \end{cases}$$

From the definition, it can be seen that p determines the amplitude and width of the non-zero part of the function, while q determines the position of the non-zero part of the function.

5.4.1.2 Haar transform matrix

The N Haar functions can be sampled at $t = m/N$, where $m = 0, \cdots, N-1$ to form an N by N matrix for discrete Haar transform. For example, when $N = 2$, we have

$$\mathbf{H}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

when $N = 4$, we have

$$\mathbf{H}_4 = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{bmatrix}$$

and when $N = 8$, we have

$$\mathbf{H}_8 = \frac{1}{\sqrt{8}} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & \sqrt{2} & -\sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{2} & -\sqrt{2} & \sqrt{2} & -\sqrt{2} \\ 2 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & -2 \end{bmatrix}$$

One important property of the Haar transform matrix is the fact that it is real and orthogonal i.e:

$$\mathbf{H} = \mathbf{H}^*, \quad \mathbf{H}^{-1} = \mathbf{H}^T, \quad \text{i.e.} \quad \mathbf{H}^T\mathbf{H} = \mathbf{I} \tag{5.23}$$

where $\mathbf{I}$ is the identity matrix.

Example 5.4.2. *When $N = 2$,*

$$\begin{aligned} \mathbf{H}_2^{-1}\mathbf{H}_2 = \mathbf{H}_2^T\mathbf{H}_2 &= \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

5.4.2 Daubechies wavelets

Daubechies wavelets defining the DWT are characterized by a maximal number of vanishing moments for the given support space. As in the Haar wavelets case the Daubechies wavelets are defined recursively, with each resolution twice that of the previous scale.

There are many Daubechies transforms, but they are all very similar. In this subsection we shall concentrate on the simplest one, the Daub4 wavelet transform. The Daub4 wavelet transform is defined in essentially the same way as the Haar wavelet transform. If a signal f has an even number N of values, then the 1-level Daub4 transform is the mapping $f \to (a^1|d^1)$ from the signal f to its first trend subsignal a^1 and first fluctuation subsignal d^1. Each value a_m of $a^1 = (a_1, \ldots, a_{N/2})$ is equal to the scalar product:

$$a_m = f \cdot V_m^1 \tag{5.24}$$

of f with a 1-level scaling signal V_m^1. Similarly, the value d_m of $d^1 = (d_1, \ldots, d_{N/2})$ is equal to the scalar product:

$$d_m = f \cdot W_m^1 \tag{5.25}$$

of f with a 1-level wavelet W_m^1. The Daub4 wavelet transform can be extended to multiple levels as many times as the signal length can be divided by 2. A signal is defined as a function that "conveys information about the behavior or attributes of some phenomenon.

The main difference between the Daub4 transform and the Haar transform lies in the way that the wavelets and scaling signals are defined. We shall first discuss the scaling signals. The scaling numbers are defined as follows:

$$\begin{aligned}
\alpha_1 &= \frac{1+\sqrt{3}}{4\sqrt{2}} \\
\alpha_2 &= \frac{3+\sqrt{3}}{4\sqrt{2}} \\
\alpha_3 &= \frac{3-\sqrt{3}}{4\sqrt{2}} \\
\alpha_4 &= \frac{1-\sqrt{3}}{4\sqrt{2}}
\end{aligned}$$

The scaling signals have unit energy. This property is due to the following identity satisfied by the scaling numbers:

$$\alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2 = 1. \tag{5.26}$$

Another identity satisfied by the scaling function is

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = \sqrt{2}. \tag{5.27}$$

(5.27) implies that each 1-level trend value $f \cdot V_m^1$ is an average of four values of f, multiplied by $\sqrt{2}$.

The Daub4 wavelet numbers are defined as follows:

$$\beta_1 = \frac{1-\sqrt{3}}{4\sqrt{2}}$$
$$\beta_2 = \frac{\sqrt{3}-3}{4\sqrt{2}}$$
$$\beta_3 = \frac{3+\sqrt{3}}{4\sqrt{2}}$$
$$\beta_4 = \frac{-1-\sqrt{3}}{4\sqrt{2}}$$

The Daub4 wavelets have unit energy. This is evident in the 1-level wavelets since

$$\beta_1^2 + \beta_2^2 + \beta_3^2 + \beta_4^2 = 1. \tag{5.28}$$

The wavelet numbers and the scaling numbers are related by the equations:

$$\alpha_1 = -\beta_4, \alpha_2 = \beta_3, \alpha_3 = -\beta_2, \alpha_4 = \beta_1.$$

Other forms of discrete wavelet transform includes the dual-tree complex wavelet transform, the undecimated wavelet transform and the Newland transform. Please refer to [183], [76] and [155] for more details of the other forms of the discrete wavelet transform.

The discrete wavelet transform has practical applications in science, engineering, mathematics and computer science.

5.5 Application of wavelets transform

In this section, we briefly discuss some real life applications of the wavelet transforms.

5.5.1 Finance

The wavelet transform has proved to be a very powerful tool for characterizing behavior, especially self-similar behavior, over a wide range of time scales. In the work of [19], the authors investigated the applicability of wavelet decomposition methods to determine if a market crash may be predictable. The premise of the work was that a predictable event produces a time series signal similar in nature with that recorded before a major earthquake which contains before-shock signals. A non predictable event would have a time series signal

resembling data produced by an explosive event which is very sudden and not preceded by any prior signals.

The wavelets methodology was very useful in yielding a correct identification of the signal type. Correct identification of signal type using wavelets techniques help mitigate some of the potential effects of the events [18]. Wavelets have also been used to investigate the dependence and interdependence between major stock markets (see [202] and [49]).

5.5.2 Modeling and forecasting

Wavelets trasnforms have been used to investigate the modeling and forecasting of nonstationary time series. [215] the authors proposes a modeling procedure that decomposes the time series as the sum of three separate components: trend, harmonic and irregular components. The proposed estimators used in their study to detect trends and hidden frequencies are strongly consistent. The ten steps ahead prediction for the time series show that the average percentage error criterion of the wavelet approach is the smallest when compared with other methods. Their results suggest that forecasting based on wavelets is a viable alternative to existing methods such as ARIMA and the seasonal ARIMA model, HoltWinters, econometric models and several others.

5.5.3 Image compression

One major application of the wavelet transform is data compression and detection of edges in digital images.

Suppose one wants to send a digital image file to a friend via email. In order to expedite the process the contents of the file needs to be compressed. There are two forms of image compression, lossless and lossy compression. Data compressed by lossless schemes can be recovered exactly without any loss of information. However for the lossy compression schemes, the data is altered. Good results can be obtained using lossy compression schemes, but the savings come at the expense of information lost from the original data ([65]).

The Haar wavelet compression is an efficient way to perform both lossless and lossy image compression. It relies on averaging and differencing values in an image matrix to produce a matrix which is sparse or nearly sparse. A sparse matrix is a matrix in which most of the elements are zero. Please refer to ([65]) for more details of image compression.

5.5.4 Seismic signals

Wavelet transforms have been applied to the decomposition of seismic signal energy into the time-scaling plane. For better representation the scale is converted into frequency. The seismic signals are obtained from hitting a metal sphere against the lawn. After a number of experiments, both the expected value of the energy center frequencies and standard deviation are found and

given in terms of the distance between the detector and place of excitation. The expected value and standard deviation decrease if the distance increases, so it is possible to estimate the distance using only one detector. These results are very useful for analyzing seismic signal whenever several excitations occur. This is demonstrated in experiments in which seismic signals are obtained from two sources. We refer to [196] for more details of the application of wavelets transform to seismic signals.

5.5.5 Damage detection in frame structures

The wavelet transform can be used to detect cracks in frame structures, such as beams and plane frames. The ability of wavelets to detect crack-like damage in structures can be demonstrated by means of several numerical examples. The method requires the knowledge of only the response of the damaged structure, i.e. no information about the original undamaged structure is required. This procedure can help detect the localization of the crack by using a response signal from static or dynamic loads. Moreover, the response needs to be obtained only at the regions where it is suspected that the damage may be present. The results of the simulation show that if a suitable wavelet is selected, the method is capable to extract damage information from the response signal in a simple, robust and reliable way. Please refer to [158] and references there in for more details.

Wavelets are known to be strong methodologies because they can capture the non-stationary behavior of the time series and are localized in both frequency and time.

5.6 Problems

1. Consider the vector u that consists of 32 equally spaced samples of the function $f(t) \approx \cos(4\pi t)$ on the interval $[0,1]$. That is $u_1 = f(0), u_2 = f(\frac{1}{32}), \ldots, u_3 = f(\frac{31}{32})$. Compute the Haar wavelet transform $y = W_{32}u$ and graph the cumulative energy u and y.

2. Define the Haar's wavelets function $\psi(t)$ and verify that for every t,

$$\psi(t) = \begin{cases} 1, & 0 \leq t < \frac{1}{2} \\ -1, & \frac{1}{2} \leq t < 1 \\ 0, & \text{Otherwise.} \end{cases}$$

3. Calculate the Haar wavelets transform for each pair s_{2k}, s_{2k+1} in the array $\overrightarrow{s} = (8,3,2,1)$.

4. Calculate the Haar wavelets transform for the data $\overrightarrow{s} = (1,3,4,1)$.

5. Discuss the similarities and differences between Fourier transform and wavelet transform.

6. Calculate the Daubechies wavelet transform of the data

$$\vec{a} = (a_0, a_1, a_2, a_3) = (2, 2, 2, 2).$$

7. For the scaling numbers,

$$\alpha_1 = \frac{1+\sqrt{3}}{4\sqrt{2}}$$
$$\alpha_2 = \frac{3+\sqrt{3}}{4\sqrt{2}}$$
$$\alpha_3 = \frac{3-\sqrt{3}}{4\sqrt{2}}$$
$$\alpha_4 = \frac{1-\sqrt{3}}{4\sqrt{2}}$$

Verify the following identity:

(a) $\alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2 = 1$

(b) $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = \sqrt{2}$

8. Show that the following eight vectors are pairwise orthogonal:

$$s1 = (1, 1, 0, 0, 0, 0, 0, 0)^T$$
$$s2 = (0, 0, 1, 1, 0, 0, 0, 0)^T$$
$$s3 = (0, 0, 0, 0, 1, 1, 0, 0)^T$$
$$s4 = (0, 0, 0, 0, 0, 0, 1, 1)^T$$
$$s5 = (1, 1, 1, 1, 0, 0, 0, 0)^T$$
$$s6 = (0, 0, 0, 0, 1, 1, 1, 1)^T$$
$$s7 = (1, 1, 1, 1, 1, 1, 1, 1)^T$$
$$s8 = (1, 1, 1, 1, 1, 1, 1, 1)^T$$

9. Haar transform

(a) For an $N \times N$ Haar transformation matrix, the Haar basis functions are

$$\psi_k(t) = \psi_{pq}(t) = \frac{1}{\sqrt{N}} \begin{cases} 2^{p/2}, & (q-1)/2^p \le t < (q-0.5)/2^p \\ -2^{p/2}, & (q-0.5)/2^p \le t < q/2^p \\ 0, & 0 \le t \le 1 \end{cases}$$

They are defined over the continuous, closed interval $t \in [0, 1]$ for $k = 0, 1, 2, \ldots, N-1$. We define the integer k such that $k = 2^p +$

$q-1$, where $0 \leq p \leq n-1$, $q = 0$ or 1 for $p = 0$, and $1 \leq q \leq 2^p$ for $p \neq 0$. The i-th row of an $N \times N$ Haar transformation matrix contains the elements of $\psi_i(t)$ for $t = 0/N, 1/n, \ldots, (N-1)/N$. Write the 5×5 Haar transformation matrix H_5 using these given conditions.

(b) Compute the Haar transfromation of the 2×2 image f with H_2:

$$f = \begin{bmatrix} 5 & -2 \\ -2 & 3 \end{bmatrix}$$

(c) The inverse Haar transform is $f = H^T T H$ where T is the Haar transform of f and H^T is the inverse of the matrix H. Show that $H_2^{-1} = H_2^T$ and use it to compute the inverse Haar transform of the result in part (a).

10. Consider the periodic signal:

$$f(t) = \begin{cases} \cos(t) + D, & -1 \leq t \leq 0 \\ \sin(t^6)/t^3, & 0 \leq t \leq 2 \end{cases}$$

where $D = \sin(2^6)/2^3 - cos(1)$, sampled at 1024 equidistant points in $[-1, 2]$. Use the wavelet transform to compress the signal. Use about 15% of the wavelet coefficients for different wavelets models. Plot the compressed signals together with the original signal using Matlab. Calculate the relative energy errors.

11. Consider the signal $f(t) = \sin(7t)$. Use Matlab to obtain the discrete wavelets transform of this signal.

12. Discuss the two main types of data (image) compression including their advantages and disadvantages.

13. Let $f(\theta)$ be the 2π-periodic function determined by the formula

$$f(\theta) = \theta^2, \quad \text{for } -\pi \leq \theta \leq \pi.$$

Find the Fourier series for $f(\theta)$.

14. Let $f(\theta)$ be the 2π-periodic function determined by the formula

$$f(\theta) = |\sin\theta|, \quad \text{for } -\pi \leq \theta \leq \pi.$$

Show that the Fourier series for f is given by

$$\frac{2}{\pi} - \frac{4}{\pi} \sum_{n=1}^{\infty} \frac{\cos 2n\theta}{4n^2 - 1}.$$

From this, show that

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{4n^2 - 1} = \frac{\pi - 2}{4}.$$