PE281 Lecture 10 Notes

James Lambers (substituting for Tara LaForce)

May 9, 2006

1 Introduction

Wavelets were developed in the 80's and 90's as an alternative to Fourier analysis of signals. Some of the main people involved in this development were Jean Morlet (a petroleum engineer), Alex Grossman, Yves Meyer, Stephane Mallat, and Ingrid Daubechies.

Waveletes are particularly useful for signal analysis, signal compression, and signal de-noising. However, they are also of use to

- geologists, for analyzing seismic data,
- FBI, for analyzing voice data
- the image processing community. The JPEG 2000 standard uses wavelets, replacing the discrete cosine transform.
- the film industry, for animation (e.g. "A Bug's Life"). Wavelets are ideal for representing changes in an image with as little data as possible, so a sequence of frames in an animation can be stored more efficiently.
- the CFD community, for solving PDE.

2 Drawbacks of Fourier Analysis

- Location information is stored in phases and difficult to extract. For example, consider discrete functions defined on an N-point grid with grid points $x_j = jh$, $j = 0, 1, \ldots, N-1$, where $h = 2\pi/N$. If $f(x)$ is defined by

$$f(x_j) = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases}$$

for some k, then the discrete Fourier transform, defined by

$$\hat{f}(\omega) = \frac{h}{\sqrt{2\pi}} \sum_{j=0}^{N-1} e^{-i\omega x_j} f(x_j),$$

is given by

$$\hat{f}(\omega) = \frac{h}{\sqrt{2\pi}} e^{-i\omega x_k}.$$

We see that the Fourier coefficients all have the same magnitude, so the only way to tell from the Fourier transform that this function is concentrated at a single point in physical space, and to determine the location of that point, is to examine the phase of the coefficients.

Because of this, it can be very difficult, if not impossible, to determine whether a signal includes a particular frequency at a particular point in physical space (which may refer to space or time). This is especially difficult for a high frequency, in view of the sampling theorem, which states that a signal with n frequencies can be represented with complete accuracy using $2n$ samples per second. However, this sampling rate must be maintained for the entire duration of the signal, not just the interval of interest.

- The Fourier transform is very sensitive to changes in the function. In view of the previous example, a change of $O(\epsilon)$ in *one point* of a discrete function can cause as much as $O(\epsilon)$ change in *every* Fourier coefficient. Similarly, a change in any one Fourier coefficient can cause a change of similar magnitude at every point in physical space.

3 Windowed Fourier Transform

To overcome these drawbacks, we could use the *Windowed Fourier Transform* (WFT), in which we take the Fourier transform of a function $f(x)$ that is multiplied by a *window function* $g(x-b)$, for some shift b called the center of the window, where $g(x)$ is a smooth function with compact support. The coefficients $F^{win}(\omega, b)$ of the WFT are given by

$$F^{win}(\omega, b) = \int_{-\infty}^{\infty} e^{-i\omega x} g(x-b) f(x)\, dx.$$

Unfortunately, this transform is difficult to invert, due to its excessive redundancy, and it does not capture short "pulses" accurately, unless a very small window is used. But in that case, low-frequency content cannot be accurately captured. This is due to the Uncertainty Principle, which states that a function cannot be simultaneously concentrated in both physical space and Fourier space.

4 Wavelets

A better solution is to use *wavelets*. A function $\psi(x)$ is a wavelet if it satisfies these conditions.

- $\int_{-\infty}^{\infty} \psi(x)\, dx = 0$
- $\int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega}\, d\omega \equiv C_\psi < \infty$

The second condition is necessary to ensure that a function can be reconstructed from a decomposition into wavelets.

5 Wavelet Families

A *wavelet family* is a collection of functions obtained by shifting and dilating the graph of a wavelet. Specifically, a wavelet family with *mother wavelet* $\psi(x)$ consists of functions $\psi_{a,b}(x)$ of the form

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right),$$

where b is the *shift* or *center* of $\psi_{a,b}$, and a is the *scale*. Alternatively, the scaling factor $1/a$ may be used. If $a > 1$, then $\psi_{a,b}$ is obtained by stretching the graph of ψ, and if $a < 1$, then the graph of ψ is contracted. The value a corresponds to the notion of frequency in Fourier analysis.

6 CWT

The *continuous wavelet transform* (CWT) of a function $f(x)$, introduced by Morlet, is defined by

$$Wf(a,b) = \int_{-\infty}^{\infty} f(x)\psi_{a,b}(x)\, dx.$$

The inverse transform is given by

$$f(x) = \frac{1}{C_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{|a|^{3/2}} Wf(a,b)\psi_{a,b}(x)\, da\, db$$

where the constant C_ψ was defined in Section 4. The CWT records *changes* in $f(x)$, which is very useful for compression or noise removal, since changes at finer scales can be omitted in a reconstruction.

7 Jump Detection

Consider the Heaviside function $H(x)$. If we use the scaling factor $1/a$ intead of $1/\sqrt{a}$ for $\psi(a, b)$, then all of the coefficients $WH(a, 0)$ of the CWT of $H(x)$ are equal. When the magnitude of the coefficients $Wf(a, b)$ do not decrease in magnitude for all a and a particular b, this suggests that there is a discontinuity in $f(x)$ near $x = b$.

8 Well-known wavelets

Some well-known wavelets are

1. Mexican hat: useful for detection in computer vision. It is the second derivative of a Gaussian function.

2. Haar: the first wavelet, introduced in 1909. It is defined by

$$\psi(x) = \begin{cases} 1 & 0 \leq x < 1/2 \\ -1 & 1/2 \leq x < 1 \\ 0 & \text{otherwise} \end{cases} .$$

   Its simple definition is helpful for computing wavelet transforms, but because it is not continuous, it is not as useful as other wavelets for analyzing continuous signals.

3. Daubechies-p: wavelets with p *vanishing moments*, to represent polynomials of degree at most $p-1$. A Daubechies-1 wavelet is equivalent to the Haar wavelet. As p increases, signals can be represented using fewer coefficients, due to fewer scales being required. On the other hand, the support of the wavelet grows with p.

9 Discrete Wavelet Transform

In practice, signals are discrete, rather than continuous. This leads to the *discrete wavelet transform* (DWT). The coefficients are defined as before, except:

1. Only particular values of a and b are used
2. Due to the discrete representation of the signal, the integrals that define the coefficients must be computed numerically.

Given a mother wavelet, an orthogonal family of wavelets can be obtained by properly choosing $a = a_0^m$ and $b = nb_0$, where m and n are integers, $a_0 > 1$ is a dilation parameter, and $b_0 > 0$ is a translation parameter.

To ensure that wavelets $\psi_{a,b}$, for fixed a, "cover" $f(x)$ in a similar manner as m increases, we choose $b_0 = \beta a_0^m$. For rapid calculation of the wavelet coefficients, we choose $\beta = 1$ and $a_0 = 2$. Note that by choosing $b_0 < 2^m$, we obtain a redundant wavelet family, whereas choosing $b_0 > 2^m$ leads to an incomplete representation of the transformed function. Therefore $b_0 = 2^m$ is the optimal choice, and in fact leads to an orthogonal family.

With these choices of a and b, the DWT of a function $f(x)$ is given by

$$Wf(m,n) = \langle \psi_{m,n}, f \rangle = \int_{-\infty}^{\infty} \psi_{m,n}(x) f(x)\, dx$$

where

$$\psi_{m,n}(x) = 2^{-m/2} \psi\left(\frac{x - n2^m}{2^m} \right).$$

The inverse transform is given by

$$f(x) = \sum_{m,n} \psi_{m,n}(x) Wf(m,n).$$

It should be noted that even though the integral defining $Wf(m,n)$ is on an unbounded interval, it is effectively on a finite interval if the mother wavelet has compact support, and therefore can easily be approximated numerically.

10 Example

We now show how the DWT of a signal can be computed, using the *Fast Wavelet Transform*, developed by Mallat. It uses two families of functions: a family of wavelets $\psi_{m,n}$, based on a mother wavelet ψ, and a family of *scaling functions* (also known as *smoothing functions*) $\phi_{m,n}$, based on a *father wavelet* ϕ. The purpose of the scaling functions is to smooth portions of the signal. The scaling and wavelet functions are used together to compute the DWT of a signal $g(x)$, by means of the following process:

1. Choose m such that the minimum spacing between samples is approximately 2^{m-1}.

2. For each integer n, compute $d_{m,n} = \langle \psi_{m,n}, g \rangle$. This extracts the highest-frequency components of g, and is referred to as applying a *high-pass filter* to g.

3. For each n, compute $s_{m,n} = \langle \phi_{m,n}, g \rangle$. This smooths the signal g, resulting in a new signal

$$\tilde{g}(x) = \sum_n s_{m,n} \phi_{m,n}(x).$$

   This step is referred to as applying a *low-pass filter* to g, as it separates the low-frequency content from the high-frequency content that has just been extracted by the high-pass filter.

4. Set $m = m + 1$ and proceed to step 2, using the smoothed signal $\tilde{g}$ in place of g.

This iteration continues until no more information about the signal can be obtained.

This entire process is called a *multiresolution analysis* (MRA) of the signal g. For each m, the set of coefficients $d_{m,n}$ describes how the signal represented by the scaling coefficients $s_{m,n}$ is changed to obtain a higher-resolution signal $s_{m-1,n}$. If we denote by V_m the vector space spanned by the scaling functions $\phi_{m,n}$, and by W_m the space spanned by the wavelet functions $\psi_{m,n}$, then the relationship between resolutions can be summarized by the relation

$$V_{j-1} = V_j \oplus W_j.$$

The direct sum is due to the fact that $\phi_{i,m}$ is orthogonal to $\psi_{j,n}$ for all i, j, m and n.

We now illustrate the FWT with a simple example. Let the signal g be a signal defined on $[0, 4)$ that is piecewise constant on each of the intervals $[x_j, x_{j+1})$, where $x_j = jh$, $j = 0, \ldots, 8$, and $h = 0.5$. The values of g on these eight subintervals are $\{0, 1, 0, 1.5, 0.5, 0, 0, 1\}$. For simplicity, we use the Haar wavelet as the mother wavelet, and the associated father wavelet, or smoothing function, is $\phi(x) = 1$ on the interval $[0, 1)$, and zero everywhere else.

One advantage of using the Haar wavelet to analyze a signal $g(x)$ is that the inner products

$$s_{m,n} = \langle \phi_{m,n}, g \rangle, \quad d_{m,n} = \langle \psi_{m,n}, g \rangle$$

can be simplified considerably, so that it is only necessary to integrate g. To see this, we consider $s_{m,n}$. Using substitutions, we obtain

$$\begin{aligned}
s_{m,n} &= \langle \phi_{m,n}, g \rangle \\
&= \int_{-\infty}^{\infty} \phi_{m,n}(x) g(x)\, dx \\
&= 2^{-m/2} \int_{-\infty}^{\infty} \phi\left(\frac{x - n2^m}{2^m} \right) g(x)\, dx \\
&= 2^{-m/2} \int_{-\infty}^{\infty} \phi\left(\frac{u}{2^m} \right) g(u + n2^m)\, du, \quad u = x - n2^m \\
&= 2^{m/2} \int_{-\infty}^{\infty} \phi\,(v)\, g((v+n)2^m)\, dv, \quad v = u/2^m \\
&= 2^{m/2} \int_{0}^{1} g((v+n)2^m)\, dv \\
&= 2^{m/2} \int_{n}^{n+1} g(w2^m)\, dw, \quad w = v + n \\
&= 2^{-m/2} \int_{n2^m}^{(n+1)2^m} g(x)\, dx, \quad x = w2^m.
\end{aligned}$$

Similarly,

$$d_{m,n} = 2^{-m/2} \left[\int_{n2^m}^{(n+1/2)2^m} g(x)\, dx - \int_{(n+1/2)2^m}^{(n+1)2^m} g(x)\, dx \right].$$

When g is piecewise constant, as in this example, these integrals can be evaluated analytically.

In computing the FWT, the first question we should address is, which scales will be needed? To answer this, we observe that g is piecewise constant on intervals of width $1/2 = 2^{-1}$, which implies that $g \in V_{-1}$. From the relations between resolutions, g can be decomposed uniquely into a signal $g_0 \in V_0$ and a *detail* signal $d_0 \in W_0$, which captures the changes that are made to g_0 to obtain g. Then, because $V_0 = V_1 \oplus W_1$, g_0 can be decomposed into a signal $g_1 \in V_1$ and a detail $d_1 \in W_1$, and so on.

Because g contains eight values, we must compute eight coefficients altogether. To see where these coefficients come from, we note that $V_{-1} = W_0 \oplus W_1 \oplus \cdots \oplus W_M \oplus V_M$, where M is the index of the coarsest scale that we will use. Therefore, the eight coefficients will come from $M + 1$ details and one smoothed signal in V_M. The detail d_0 contains four coefficients, because the interval $[0, 4)$ is covered by the four wavelet functions $\psi_{0,n}$ for $n = 0, 1, 2, 3$. The next detail d_1 contains two coefficients, and the next one,

d_2, contains one, for a total of seven. The eighth coefficient is obtained from the smoothed signal g_2. Therefore, $M = 2$ and three scales are needed. In general, for a signal with N points, $\log_2 N$ scales are needed to represent it completely.

We begin with $m = 0$ and apply the high-pass filter, computing

$$\begin{aligned} d_{0,n} &= \langle \psi_{0,n}, g \rangle \\ &= \int_{-\infty}^{\infty} \psi_{0,n}(x) g(x)\, dx \\ &= \int_{-\infty}^{\infty} \psi(x-n) g(x)\, dx \\ &= \int_{n}^{n+1/2} g(x)\, dx - \int_{n+1/2}^{n+1} g(x)\, dx, \quad n = 0, 1, 2, 3. \end{aligned}$$

Because ϕ is piecewise constant, we can easily evaluate these integrals exactly, and obtain

$$d_{0,0} = -0.5, \quad d_{0,1} = -0.75, \quad d_{0,2} = 0.25, \quad d_{0,3} = -0.5.$$

Next, we apply the low-pass filter, computing

$$\begin{aligned} s_{0,n} &= \langle \phi_{0,n}, g \rangle \\ &= \int_{-\infty}^{\infty} \phi_{0,n}(x) g(x)\, dx \\ &= \int_{-\infty}^{\infty} \phi(x-n) g(x)\, dx \\ &= \int_{n}^{n+1} g(x)\, dx, \quad n = 0, 1, 2, 3. \end{aligned}$$

We obtain

$$s_{0,0} = 0.5, \quad s_{0,1} = 0.75, \quad s_{0,2} = 0.25, \quad s_{0,3} = 0.5.$$

It follows from the relation between resolutions, $V_{-1} = V_0 \oplus W_0$, that g has a unique decomposition into the sum of an element from V_0 and an element of W_0. Specifically,

$$g(x) = \sum_{n=0}^{3} s_{0,n} \phi_{0,n}(x) + \sum_{n=0}^{3} d_{0,n} \psi_{0,n}(x),$$

where the first sum is a smoothed, lower-resolution version of g, and the second sum represents the changes made to this smoothed version to obtain g.

We now continue the transform. Working with the smoothed signal

$$g_0 = \sum_{n=0}^{3} \phi_{0,n} s_{0,n} = \{0.5, 0.75, 0.25, 0.5\} \in V_0,$$

we apply the high-pass filter corresponding to $m = 1$, computing

$$\begin{aligned} d_{1,n} &= \langle \psi_{1,n}, g_0 \rangle \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2}} \psi\left(\frac{x-2n}{2}\right) g_0(x)\, dx \\ &= \frac{1}{\sqrt{2}} \int_{2n}^{2n+1} g(x)\, dx - \frac{1}{\sqrt{2}} \int_{2n+1}^{2n+2} g(x)\, dx, \quad n = 0, 1. \end{aligned}$$

We obtain

$$d_{1,0} = \frac{-0.25}{\sqrt{2}}, \quad d_{1,1} = \frac{-0.25}{\sqrt{2}}.$$

Next, we apply the low-pass filter to obtain

$$s_{1,0} = \frac{1}{\sqrt{2}} \int_0^2 g(x)\, dx = \frac{1.25}{\sqrt{2}}, \quad s_{1,1} = \frac{1}{\sqrt{2}} \int_2^4 g(x)\, dx = \frac{0.75}{\sqrt{2}}.$$

We now have

$$g(x) = \sum_{n=0}^{1} s_{1,n} \phi_{1,n}(x) + \sum_{n=0}^{1} d_{1,n} \psi_{1,n}(x) + \sum_{n=0}^{3} d_{0,n} \psi_{0,n}(x),$$

where the first sum is an even smoother approximation to g than g_0, the second sum describes the high-resolution changes made to this smoother signal to obtain g_0, and the third sum describes the higher-resolution detail added to g_0 to obtain g.

We continue with $m = 2$, and the smoothed signal $g_1 = \{0.625, 0.375\}$. The high-pass filter yields

$$\begin{aligned} d_{2,0} &= \langle \psi_{2,0}, g_1 \rangle \\ &= \int_{-\infty}^{\infty} \psi_{2,0}(x)\, g_1(x)\, dx \\ &= \int_{-\infty}^{\infty} \frac{1}{2} \psi\left(\frac{x}{4}\right) g_1(x)\, dx \\ &= \frac{1}{2} \int_0^2 g(x)\, dx - \frac{1}{2} \int_2^4 g(x)\, dx = 0.25, \end{aligned}$$

while the low-pass filter yields

$$\begin{aligned} s_{2,0} &= \langle \phi_{2,0}, g_1 \rangle \\ &= \int_{-\infty}^{\infty} \phi_{2,0}(x)\, g_1(x)\, dx \\ &= \int_{-\infty}^{\infty} \frac{1}{2}\phi\left(\frac{x}{4}\right) g_1(x)\, dx \\ &= \frac{1}{2}\int_0^4 g(x)\, dx = 1. \end{aligned}$$

This corresponds to the signal $g_2(x) = s_{2,0}\phi_{2,0}(x) = 0.5$ that is constant on $[0, 4)$.

The signal g_2 can be represented by a single wavelet function in W_3 provided a suitable extension for g_2 beyond the interval $[0, 4)$ is used. If we extend the signal g_2 so that it is equal to -0.5 on the interval $[4, 8)$, then we can apply the high-pass filter one more time and obtain

$$d_{3,0} = \int_{-\infty}^{\infty} \frac{1}{2\sqrt{2}}\psi\left(\frac{x}{8}\right) g_2(x)\, dx = \sqrt{2}$$

while the low-pass filter yields

$$s_{3,0} = \int_{-\infty}^{\infty} \frac{1}{2\sqrt{2}}\phi\left(\frac{x}{8}\right) g_2(x)\, dx = 0.$$

We conclude that with the extension to $[4, 8)$, $g_2 \in W_3$, and therefore $g \in W_0 \oplus W_1 \oplus W_2 \oplus W_3$, and can be expressed as

$$g(x) = \sum_{m=0}^{3} \sum_{n=-\infty}^{\infty} d_{m,n}\psi_{m,n}(x),$$

using the inverse DWT.

11 Compression and De-Noising

As the previous example shows, a discrete signal with N samples can be represented by N members of a wavelet family. However, if the detail coefficients corresponding to finer scales are negligible, they can be omitted from the MRA, allowing the signal to be compressed without loss of accuracy. Also, such coefficients may represent the noise in a signal, so dropping them has the effect of de-noising the signal.