Introduction to Wavelets

Lecture #1: Tuesday, 30 September 1997
Lecturer: Denis Zorin
Scribe: John Owens

1 Why are wavelets useful?

Wavelets have a few interesting applications, some of which are mentioned below. However, the applications of wavelets by themselves are limited. The ideas behind wavelets, which we will be covering in this lecture and future lectures, are more important.

The most common use of wavelets is in signal processing applications. For example:

- Compression applications. If we can create a suitable representation of a signal, we can discard the "least significant" pieces of that representation and thus keep the original signal largely intact. This requires a transformation which separates the "important" parts of the signal from less important parts.

  In the simplest case, we can decompose a signal into two parts: a low frequency part, which is some sort of average of the original signal, and a high frequency part, which is what remains after the low frequency part is subtracted from the original signal. If we are interested in the low frequency part and hence discard the high frequency part, what remains is a smoother representation of the original signal with its low frequency components intact. Alternatively, if we are most interested in the high frequency part, we may be able to discard the low frequency part instead.

  This approach, that of decomposing a signal into two parts, is common for all wavelets. Also fundamental to wavelet analysis is a heirarchical decomposition, in which we may apply further transforms to an already decomposed signal.

- Edge detection. With this application it is most important to identify the areas in which the input image changes quickly. We can discard the smooth (low frequency) parts. The simplest wavelet basis, the Haar basis (to be discussed later) is suitable for this application.

  Along this vein, the book by Strang and Nguyen describes a widely used application of wavelets, fingerprint compression, in which edge detection figures prominently.

- Graphics. Two prominent uses of wavelets in graphics include

1. Curve and surface representations; and
2. Wavelet radiosity.

These two reflect two quite different uses of wavelets.

- Numerical analysis. Wavelets are used in the solution of partial differential equations and integral equations.

2 History of Wavelets

The first use of wavelets was by Haar in 1909. He was interested in finding a basis on a functional space similar to Fourier's basis in frequency space. In physics, wavelets were used in the characterization of Brownian motion. This work led to some of the ideas used to construct wavelet bases. Wavelets were also used for analysis of coherent states of a particular quantum system. Finally, in the signal processing field, S. Mallat discovered that filter banks have important connections with wavelet basis functions.

3 Filters and Filter Banks

3.1 Linearity and Time Invariance

Consider a discrete input signal $x(n)$, a filter H, and an output $y(n)$. We express the operation of H on the input signal x as $y = Hx$.

We call the filter H "linear" if scaling the input scales the output, and we call H "time invariant" if shifting the input (in time) correspondingly shifts the output. In these notes all filters will be assumed to be linear and time invariant.

3.2 Filter operation

If H is linear and time invariant, the we can express its operation as follows:

$$y(n) = \sum_k h(k)x(n-k).$$

This operation is called a convolution. The individual coefficients $h(i)$ are the "impulse responses" of the system.

The equation $y = Hx$ can also be written as an infinite matrix, with y and x column vectors and H in the following form:

$$\begin{bmatrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & h(1) & h(0) & h(-1) & \cdots & \cdots \\ \cdots & \cdots & h(1) & h(0) & h(-1) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

Note that the entries of the filter are in reverse indexed order. If the input stream is finite with n elements, H is a finite $n \times n$ matrix.

3.3 What is a basis?

For a given space of functions, a frame is a collection of functions such that any function in the space is a weighted sum of the functions in the frame. In other words, if the functions in the frame are $f_0 \ldots f_n \ldots$, then any function g can be written as $g = \sum_i a_i f_i$.

A basis is also a collection of functions. Any basis is a frame, but a basis also has the property of linear independence. With a basis, the coefficients a_i of the expansion of the function g (written as a weighted sum of the basis functions $g = \sum_n a_i f_i$) are uniquely determined. Another way to state this property is that no function f_i in the basis can be written as a weighted sum of the other functions of the basis.

For a function space we can define an inner product. One example of an inner product is a dot product, used in a vector space. In a function space we define the inner product $< f, g >$ of two functions f and g as $\int f(t)g(t)dt$.

One desirable property of a basis is orthogonality. With an orthogonal basis, the inner product of two basis functions f_i and f_j is equal to zero if $i \neq j$.

A second desirable property is orthonormality, which implies that taking the inner product of a basis function f_i with itself equals 1. For our functional basis we see that $\int f^2(t)dt = 1$.

In summary, an orthogonal, orthonormal basis implies that

$$< f_i, f_j >= \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

So why is orthonormality a desired property? Let us define a function x as the weighted sum of the basis functions, $x = \sum a_i f_i$. We would like to find the coefficients a_i, and orthonormality makes this simple. We only need to take the inner product of the function x with a_i's associated basis function to find the associated coefficient a_i.

$$\begin{aligned} < x, f_j > &= \sum_i a_i < f_i, f_j > \\ &= a_j < f_j, f_j > \\ &= a_j \end{aligned}$$

3.4 Filters and wavelets, Haar filters

Each wavelet basis has two filters associated with it. In general those filters are expressed in the form $y(n) = \sum h(n)x(n)$.

The Haar basis is particularly simple. The two filters are H_0 and H_1 and are defined as follows:

$$H_0 : y(n) = \frac{1}{2}x(n) + \frac{1}{2}x(n-1)$$

$$H_1 : y(n) = \frac{1}{2}x(n) - \frac{1}{2}x(n-1)$$

For H_0, all h coefficients are zero except for $h_0(0) = 1/2$ and $h_0(1) = 1/2$. Similarly, in H_1, $h_1(0) = 1/2$ and $h_1(1) = -1/2$.

H_0 computes a moving average of its input, resulting in a sequence which is smoother than the initial sequence. Hence it is a low pass filter. H_1 computes a moving difference and serves as a high pass filter.

3.5 Scaling functions and the dilation equation

We define a function ϕ as a "scaling function". Scaling functions obey the dilation equation,

$$\phi(t) = 2\sum h_0(k)\phi(2t-k). \tag{1}$$

Note that the filter coefficients used (the h_0's) are from the first of the pair of filters. We will use the second filter later.

If this equation has an appropriate solution $\phi(t)$, then we can construct a *frame* using any integral values of j or k in the following generating relation:

$$\phi(2^j t - k). \tag{2}$$

Next we introduce the wavelet equation (3), which uses the coefficients from the second pair of filters:

$$w(t) = 2\sum h_1(k)\phi(2t-k). \tag{3}$$

We generate a wavelet *basis* with a similar generating relation

$$w(2^j t - k).$$

Note that scaling functions don't form a basis; wavelets do. But wavelets are constructed from scaling functions, so we have to find a scaling function first.

As can be seen from the wavelet equation, wavelets are defined as linear combinations between scaling functions one level below, where a "level" is defined as the set of all scaling functions generated with a given j.

For instance, the Haar scaling function satisfies the dilation equation with the solution

$$\phi(t) = \phi(2t) + \phi(2t-1).$$

And the Haar wavelet satisfies the wavelet equation with the solution

$$w(t) = \phi(2t) - \phi(2t-1).$$

The Haar basis spans the space of all functions with integrable square.

Haar wavelets are orthogonal.

4 Filters and Filter Banks

4.1 Definition of several operators

Here we will define four operators and describe their operation (Figure 1).

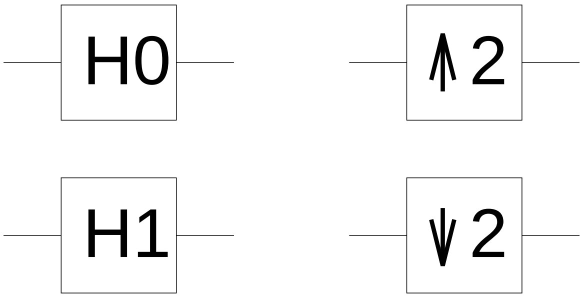

Figure 1: Our 4 operators

The first operator, $H = H_0$, is the Haar moving average filter. It outputs the average of its current input and its previous input.

The second operator, H_1, is the Haar moving difference filter. It outputs half the difference between its current input and its previous input.

The third operator, $\uparrow 2$, is an upsampling operator. It outputs each input twice, outputting at double the rate of its input.

The fourth operator, $\downarrow 2$, is a downsampling operator. It outputs every other input it receives, outputting at half the rate of the input.

4.2 Overview

We will look at two filter banks, an analysis bank and a synthesis bank. The analysis bank will use the Haar filter pair, consisting of a low pass and a high pass filter. We would like to pick a synthesis filter bank such that the output of the analysis and synthesis filter connected in series is the same as the input into the analysis filter, with perhaps a time delay.

First we will look at the analysis bank, then the synthesis bank, then we will put them together and show that the aforementioned property holds.

4.3 Haar analysis bank

Figure 2 shows the analysis bank.

The input comes in on the left side, and the output of the filter bank leaves through the two paths on the right side. It is desirable (but not necessary) that the bandwidth and storage requirements in the system be equal to those of the input stream. Hence, the downsampling operators cut the bandwidths in half (compared to that of the input

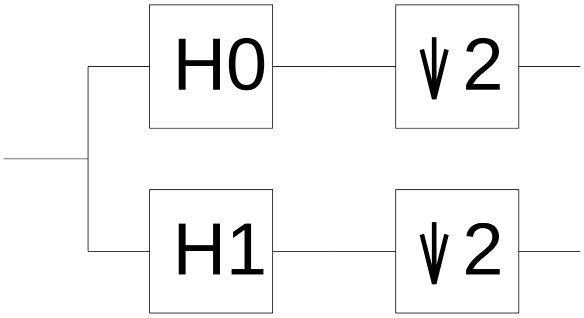

Figure 2: The Haar analysis bank

stream) on each of the two output streams, and the overall bandwidth requirement entering the analysis bank is equal to that leaving the bank. In this case the filter bank can be viewed as a linear transform: if the input signal had finite length n, the output will also have length n, consisting of two halves, each having length $n/2$.

4.4 Synthesis bank

Given the above analysis bank, what synthesis bank should we use to reproduce the input on the output? The basic analysis filter structure is shown in Figure 3. The output of each of the upsampling filters has the same bandwidth as the initial input stream, but note the outputs of filters F and G are summed to create an output stream with the same bandwidth as the original input stream.

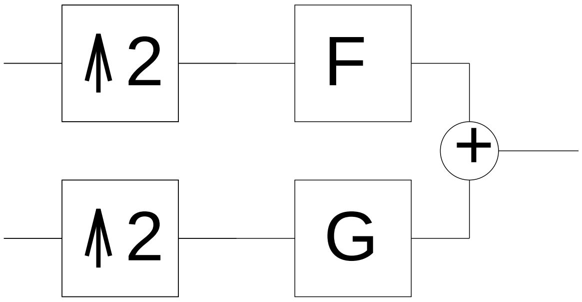

Figure 3: The basic synthesis bank

Conveniently enough, it turns out that we can use the Haar moving average filter H_0 as the filter F in Figure 3 and similarly, replace G with $-H_1$. This structure is shown in Figure 4.

The complete filtering system is shown in Figure 5.

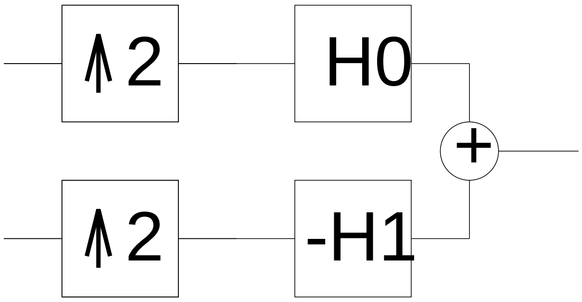

Figure 4: The Haar synthesis bank

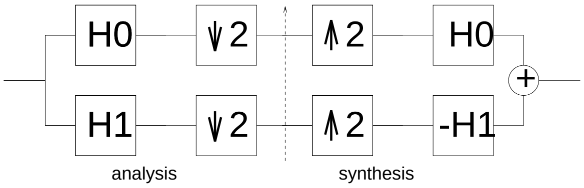

Figure 5: The complete system

5 Transforms used for analysis and construction of filters

Two related transforms are the Fourier transform and the z transform. To see more background on Fourier transforms, please consult the handout on "Antialiased Shifting and Resizing" distributed in class. Both transforms are described in detail in most signal processing books.

5.1 The Fourier Transform

The Fourier transform is defined on both a discrete signal $x(k)$ as

$$X(\omega) = \sum_k x(k) e^{-jk\omega} \tag{4}$$

and a continuous signal $x(t)$ as

$$X(\omega) = \int x(t) e^{-jk\omega} dt.$$

Following the signal processing tradition, we use j to denote $\sqrt{-1}$.

5.2 The z Transform

The z transform is similar but does not use powers of $e^{-j\omega}$. instead choosing powers of z:

$$X(z) = \sum_k x(k) z^{-k}.$$

5.3 Frequency response

The frequency response of a filter H is written as $H(\omega)$ (or $H(e^{j\omega})$). In Fourier space, the frequency response is useful for analyzing linear time invariant filters. Why? Convolution (i.e. the operation we use to describe filter operation) is simply multiplication. In other words, the operation

$$y(n) = \sum_k h(k)x(n-k)$$

in Fourier space is equivalent to

$$Y(\omega) = H(\omega)X(\omega).$$

We can see, then, that $H(\omega)$, used as a filter, serves to emphasize or deemphasize frequencies in the input signal. Each frequency component of the input $X(\omega)$ corresponding to a frequency ω is scaled by $H(\omega)$. For example, a low-pass filter passes through components with small ω ($H(\omega)$ is close to 1 when ω is small) and suppresses high-frequency components ($H(\omega)$ is close to 0 if ω is close to π).

5.4 Frequency response of the Haar basis

First we will look at the frequency response of H_0, the moving average function. Recall that the frequency response of an input stream is defined in (4). In H_0 the only coefficients that are non-zero are $h(0) = h(1) = 1/2$.

Thus we can calculate the input response using (4):

$$\begin{aligned} H_0(\omega) &= (1 + e^{-jk\omega})/2 \\ &= e^{-jk\omega/2}(e^{jk\omega/2} + e^{-jk\omega/2})/2 \\ &= e^{-jk\omega/2}\cos(\omega/2) \end{aligned}$$

and $|H_0(\omega)| = \cos(\omega/2)$.

With a similar derivation we can show that $|H_1(\omega)| = |\sin(\omega/2)|$.